Simulation of Asset Prices using Lévy Processes

Mark L. Riemer
Department of Mathematics and Applied Mathematics
University of Cape Town

supervised by
Dr. Peter Ouwehand

Dissertation

presented to the Faculty of Science
of the University of Cape Town

in partial fulfillment of the degree of
M.Sc. in the Mathematics of Finance

November 25, 2008

Abstract

Pricing financial instruments is central to modern Finance, and a heavy reliance on the use of Mathematical techniques has been used to price particularly the more exotic style and over-the-counter derivatives. There has been movement away from the traditional geometric Brownian motion models, for two main reasons: volatility is not constant, and price processes exhibit jumps. An additional aspect of modelling that has become quite important is the use of Monte Carlo techniques, and these require simulation from the sample paths of asset price processes, as do options which have a (possibly complex) path dependence.

This dissertation focuses on a Lévy process driven framework for the pricing of financial instruments. The main focus of this dissertation is not, however, to price these instruments; the main focus is simulation based. Simulation is a key issue under Monte Carlo pricing and risk-neutral valuation – it is the first step towards pricing and therefore must be done accurately and with care. This dissertation looks at different kinds of Lévy processes and the various approaches one can take when simulating them.

Contents

Chapter 1

Modelling beyond Brownian Motion

1.1 Introduction

In recent years, mathematics has enjoyed considerable success in applying itself to the vast and complex world of financial modelling. It is a relatively young field, but has quickly become one with a high level of actively researched topics and ideas. Using quantitative techniques for the pricing of assets through the modelling of their price movements is, these days, common practice throughout the derivatives market. A careful and consistent pricing framework is required so that buyers and sellers of these instruments are not exposed to arbitrage. Markets, generally speaking, are assumed to be arbitrage-free. However, this is not a claim which is easy to verify. Put into weaker form, it states that arbitrage opportunities evanesce as arbitrageurs exploit mispricings to make profits. Their exploitations affect supply and demand which in turn affects the price, causing it to move towards an equilibrium price, in line with arbitrage pricing theory (APT).

Today, a substantial amount of research has gone into models of asset dynamics – these include stock prices and interest or exchange rates. The use of *stochastic processes* for financial modelling has a significant history: Random walks, diffusions or – as they are more commonly known – Brownian motions (with drift) were initially proposed by Bachelier in 1900 in arithmetic form as a model for stocks on the Paris Bourse. These, however, received little to no significant attention until the 1960's. One problem with Bachelier's model was that it accommodated negative asset prices. It was only 60 or so years later that the idea of using a geometric Brownian motion was introduced by Paul Samuelson [73] in 1965. This was a forerunner to the the ground-breaking 1973 paper by Black and Scholes [15] which gave the market a formula still widely used in current markets, albeit with a lot more skepticism – the Black-Scholes formula. Brownian motion is not the only process which can be used in a context of financial modelling. The inherent properties of a Brownian motion can also be found in a more general class of processes. These are known as Lévy processes, named after the French mathematician Paul Lévy . By studying their properties, ways and means of simulating from their associated distributions can be developed. This means that they can be incorporated into derivative pricing models. Lévy processes were introduced in this context as a model for cotton prices by Mandelbrot [60] in 1963.

The popularity of the cornerstone of modern finance, the Black-Scholes model is due, in part, to its provision of simple, closed form solutions for arbitrage-free prices of vanilla puts and calls. It has been shown to be consistent with binomial tree option pricing [46, 58]: As the number of time steps in a binomial tree pricing scheme is increased arbitrarily over a finite time horizon, the prices of calls (and puts) can be shown to converge to the very prices given by the Black-Scholes

formula. Since the underlying theory is based on an assumption that asset returns[1] are normally distributed, simulation of the underlying asset dynamics is straightforward. The substantial literature on efficient simulation techniques of Gaussian random variables ([17, 27, 64] for example) makes the task of pricing instruments with complex, path-dependent, payoff structures (exotic options) much more tractable. The advent of Monte Carlo and Quasi-Monte Carlo pricing techniques assist in this regard. In reality, these days, prices for vanilla puts and calls are quoted by the market. This means that they are not really "derivatives" anymore, since derivatives are defined as instruments which derive their price from some underlying asset, and that 'fair' price is determined by some form of arbitrage argument. What has happened instead is that vanilla options have become *primary instruments* on markets (not all that different from their equity counterparts) – their prices are not determined by arbitrage arguments. This is not to say that the notion of arbitrage in the prices has been defenestrated, because arbitrage arguments work both ways. A large trading activity in stock options, would incur price movements of that option and should be partially responsible for associated price movements in the underlying asset in order to eliminate induced arbitrage opportunities.

Even these days, market models are still dominated by those with a strong dependence on the Black-Scholes-Merton framework. Variations of these models include those which accommodate stochastic volatility (Hull & White, 1987 [47]; Dupire, 1994 [28] for example) or non-constant interest rates and models of the bond market (Vasicek, 1977 [81] ; HJM 1992 models [43]; Musiela-Rutkowski, 1997 [69] for example). The option of considering things from a different perspective altogether has only very recently become popular. This approach, out of criticism for the assumptions of the Black-Scholes model and its results, suggests that an alternate framework is required. The very need for these 'extension' models mentioned above stems from a subtle acknowledgement that the Black-Scholes model is unable to explain certain crucial features of asset returns. One aim of these models is to introduce greater flexibility into the model through an increase of the number of parameters. The issues raised by cynicism regarding this seminal model are worth further investigation.

1.2 Empirical properties of asset returns

A crucial assumption of the Black-Scholes model is that underlying asset log-returns (at maturity) are normally distributed [15, 46]. Modelling this imposes two requirements: stock volatility must be constant and price (sample) paths must be continuous. A violation of one of these conditions is an indication of departure from normality in the log-return distribution. There is a growing body of documented evidence treating the claim of normally distributed log-returns with suspicion, dating as far back as 1965, to a paper by Fama [33]. Other sources include Madan (1999) [58], Geman, Madan & Yor (1998) [38] and Geman (2002) [37]. Generally, today, this assumption is refuted – simple estimation of any arbitrary assets return distribution will readily suggest this. A brief synopsis of empirically observed evidence of asset return characteristics may help to motivate an alternative approach of modelling asset returns through Lévy processes. These are taken from a discussion in Cont & Tankov [25]:

1. **Heavy tails:** Actual share returns appear to have greater probabilities associated with larger movements, having an unlikely associated probability under the normal distribution. This heavy tail phenomenon is particularly pronounced for high frequency returns and short term data [58]. The tail behaviour of asset returns has been studied by certain people (Jansen and de Vries, 1991 [51]; Longin, 1996 [57]; Dacorogna *et al*, 1998 [26] for

[1] logreturns

example) and as mentioned in [25] the empirical findings seem to be that stocks, market indices and exchange rates appear to have tails which have an index of between 2 and 5. There is a class of Lévy processes known as α-stable processes and these have a tail index $\alpha \in (0, 2)$. These authors conclude that tails are observed to be generally lighter than the tails of α-stable Lévy processes and that they have finite variance. The Cauchy distribution (associated with a Lévy process in the class of general hyperbolic (GH) Lévy processes) has infinite variance while the variance-gamma or normal inverse Gaussian models have infinite tail index. These properties weaken arguments for their use as effective models in this case. On the other hand the Student-t distribution, it can be shown, gives rise to a GH Lévy process which may be a better process to use, as it has finite tail index and variance [25].

2. **Gain/loss asymmetry (skewness):** For equity data, large negative returns – it is argued – are more likely than their positive counterparts which translates into a fatter left tail. This effect appears to be less pronounced in exchange rate data - there is more of a symmetry in the tails as larger negative movements are as likely as large positive movements. This implies that the tails of the distribution are more or less equally fat. As with heavy tails, the effect is more pronounced for high frequency returns in the short term.

3. **Absence of autocorrelations:** Linear autocorrelations are often insignificant, unless one is dealing specifically with small intraday returns (of the order of 20 minutes). Here there are probably some microstructure effects which come into play.

4. **Aggregational normality:** As the time scale for the calculations of returns is increased, their distribution appears to tend towards a normal distribution.

5. **Volatility clustering:** Large price changes appear to be followed by large price changes, while small price changes tend to follow on from small price changes. This is quantitatively manifested by noting that returns, $r(t)$, are uncorrelated, but their squares, $|r(t)|^2$, display a slowly decreasing, positive and significant autocorrelation function. This motivated the use of GARCH modelling for time-series of returns.

6. **Conditional heavy tails:** Even after correcting for volatility clustering using GARCH or somesuch, the residual time series still appears to exhibit heavy tails. The tails appear to be less heavy than the unconditional tail distribution of the returns.

7. **Slow decay of autocorrelation in absolute returns:** The squared returns' autocorrelation/time lag function appears to decay slowly, as some sort of power law. This can sometimes be interpreted as a sign of long-range volatility dependence.

8. **"Leverage" effect:** As a company's equity value declines its leverage increases and the asset becomes more of a risky investment - hence the volatility increases. On the other hand, as the company's equity value increases, then its leverage decreases and so does the associated volatility. The negative correlation between returns of an asset and its volatility is termed the 'leverage effect' and is one way to account for the non-constancy of equity volatility [46]. Heuristically, the proposal of constant volatility is a bit far-fetched and idealistic. In reality, price movements affect the overall confidence of the market: Falling stock prices feed back into the market as a fear that they will continue to fall which increases underlying volatility. When prices rise, the expectation that they are on a 'good run' is interpreted as a decrease in the underlying volatility.

9. **Volume/volatility correlation:** There seems to exist a positive correlation between this (and other measures of market activity) with all measures of volatility.

10. **Asymmetry of time scales:** So called "coarse-grained" measures of volatility predict fine-scale volatility much better than the other way around.

11. **Self-similarity of asset returns:** Whether asset returns are self-similar or not is also a point worth considering. The reason for this is that Brownian motions are self-similar – a property allowing the same model to be used for options across different maturities, or different time-scales. Self-similar Lévy processes include Brownian motions[2] and the so called stable processes (with index of stability, α) [25]. Rejecting self-similarity means the use of an α-stable Lévy process for modelling log-returns must be rejected, thus forfeiting the associated 'time-horizon flexibility' property. It is perhaps best argued that asset returns are not strictly self-similar – their distributional properties (variance, kurtosis, skewness amongst others) change as the time interval, Δ, used in the calculation of asset returns, changes. Studies, mentioned in [25], (Akgiray & Booth [5] (1988); Blattberg & Gonedes (1974) [16]), show that the value of α seems to increase as the time horizon increases. This may also impact the choice of Lévy process used for modelling.

Disagreement and subjective views on which of the above properties are most important will influence the construction of an asset return model. Lévy processes are not able to account for, or explain away, all of the above observed phenomena, but they do seem to naturally incorporate more of these properties into their modelling than straightforward geometric Brownian motions. Certain Lévy processes are able to explain specific properties better than others, motivating the use of different Lévy processes depending on the task at hand.

1.3 Problems with the Black-Scholes framework

Today, the implications of many of these empirical and statistical findings manifest themselves in what has become known worldwide as the implied volatility smile or skew. Generally the word "smile" is used to refer to the curvature of the volatility/strike function while "skew" refers to its slope. It exists as a testament to the fact that output from the formula of Black and Scholes requires modification. The smile became the new best friend of traders and fund managers after the fabled Wall Street crash of 1987 - particularly for equity options. The smile is particularly well known and used in foreign exchange options markets and equity options markets, although the smile typically exhibits different shapes in the different markets [46].

In addition to this, because vanilla instruments are *primary instruments*, under the Black-Scholes framework traders and quants have to get used to extensively working with, a quantity dubiously defined as the *implied volatility* - which actually has nothing to do with the traditional meanings of the word 'volatility'. It is simply the value for σ that must be shoved into the Black-Scholes formula in order to obtain the correct, quoted option price. Pricing vanilla options is no longer an issue, however the pricing of exotic options is generally of interest. This *has to* happen in a manner consistent with the the pricing of primary instruments, so as to protect against arbitrage. Herein lies the extreme importance of the use of the market *volatility smile*. There are many (some good, some bad) calibration models in the literature (see [22] or [83] for example) which fit the skew to market data, but the primary reason behind the existence of a skew is because the Black-Scholes model is not perfect.

[2] although Brownian motions with drift are not self-similar, they are known as self-affine processes

Generally, the lognormal claim is refuted: The market does not believe that asset log-returns[3] are normally distributed. The assumption, either, that discontinuities are present in price paths or that volatility is not constant invalidates the lognormal assumption. A remarkable theorem due to Breeden & Litzenberger (1976) [18] allows one to determine the relationship between the empirical return distribution and the volatility skew: if the return distribution is indeed lognormal, then volatility is a constant (as expected). However, if the implied distribution has fatter tails than the lognormal density, this translates into a higher volatility – on either side – than that of the (constant) lognormal implied volatility. The volatility will be shaped like a smile, hence the name. A thinner right hand tail implies that the volatility in that region is lower than the lognormal implied volatility, also producing a smile shape (perhaps more of a smirk). One motive for using Lévy processes, in general, claimed in Madan (1999) is that it is intentionally possible to flatten the smile by use of a suitable Lévy process (in their case, a Variance-Gamma process) [58]. So for a given skew, it is possible to back out an implied risk-neutral distribution allowing risk-neutral valuation or pricing to take place.

A further potential pitfall of working in the Black-Scholes world is that the underlying stochastic process used for modelling is based on a (continuous) Brownian motion or random walk[15]. As a mathematical entity, Brownian motion has some remarkably awkward properties, which are perhaps not all that ideal for modelling asset prices or returns. Brownian motions, by definition and construction, have continuous sample paths (see Hunt & Kennedy [48], Sato [74] for example). This is a limitation which has the effect of discarding (with negligent probability) large price movements - or jumps. Any model based on a Brownian motion (whether arithmetic or geometric) will therefore lack the ability to explain any 'jumps' present in asset price processes. Jumps are empirically present in arbitrary asset returns, and should therefore be incorporated into any kind of model aiming to model asset returns properly [25, 79]. One way to achieve this is through the use of Lévy processes. As already mentioned, Brownian motion (with drift), is a Lévy process and is the only such process which is continuous[37]. Arguing against using continuous price processes can be more subtly motivated by noting that the inability to trade continuously should have a *de facto* correspondence to the claim that price processes and their movements are not entirely of a continuous nature [37].

Modelling with a continuous diffusion process implies that information arrives all the time (at a continuous rate) and that the asset price adjusts accordingly, under the theory of Efficient Markets. Intuitively, there is reason enough to cast doubt on this claim. Buyers and sellers interact in the market in a random fashion themselves – busy patches and quiet patches are not uncommmon along with block trades and other things. An inherent stochasticity on event times is implied by these occurrences – perhaps modelled by the exponential (or gamma distributions) due to their conventional use for modelling inter-arrival times. Their use was initially motivated because of the "lack of memory property" which is exhibited by the exponential distribution. The same is true for the arrival of new (and exciting, price-adjustable) information. The restriction of continuous time modelling can be circumnavigated by using an alternate Lévy process.

Sample paths of Brownian motions suffer from being of infinite variation on any bounded interval: A Brownian motion starting at 0, will cross the x-axis infinitely often – on any bounded interval $[0, u]$ where $u > 0$. Before the variation of a Brownian motion becomes finite, the absolute changes must first be squared before they are added. Thus, a Brownian motion has finite quadratic variation but infinite variation. Specific asset prices (such as equity indices), it is motivated, are better represented by finite variation processes [37].

[3] continuously compounded returns are widely used, and are defined as $r_t(\Delta) = \ln(P_{t+\Delta}/P_t)$. Thus, claiming that returns are lognormally distributed and log-returns are normally distributed are in fact identical

In general, markets are assumed to be incomplete. If markets were complete, it would be possible to replicate every contingent claim by a self-financing portfolio. This reduces the need for derivatives entirely [25]. The primary motivation for using continuous processes to represent price movements did not stem from a desire for accuracy, but rather because of a dynamic hedging argument, which was valid in that context, but essentially made options redundant instruments [38]. Geman, Madan and Yor (2002) claim in [38] that the use of discontinous processes implies options are no longer replicable by simple trading in the stock and money market accounts (as in the Black, Scholes and Merton framework). In addition to this, with discontinuous paths they obtained greater accuracy. Here, options are seen as market *completing* assets having a worthwhile role to play. The options themselves are useful in hedging jump risks, while the option prices constitute a rich source of information which can be employed in the design of optimal risk measures.

1.4 Using Lévy processes

The use of the geometric Brownian motion model for modelling asset prices can be argued economically: The time series process of an assets price return series should be stationary, and any shocks which occur should be independent. The central limit theorem then suggests that these returns should be normally distributed, hence the use of the geometric Brownian motion model. This model, however, does not work well, so one approach to finding a better model is then to weaken a non-economic requirement, such as that of path-continuity. The implication is that models based on Lévy processes must be used. The definition of a Lévy process is similar to that of a Brownian motion (with drift), with the exception that the condition of pathwise continuity is weakened to that of continuity in probability. That is: a Lévy process is a process with stationary independent increments such that $X_s \to X_t$ in probability whenever $s \uparrow t$. It can be shown that the only continuous Lévy processes are Brownian motions [25]. A formal definition is given in Chapter 2. There are many different types of Lévy processes defined through the literature and these will be further investigated in Chapter 3.

A class of statistical distributions important in the theory of Lévy processes is the class of *infinitely divisible* distributions. This is because there is a remarkable relationship between Lévy processes and infinitely divisible distributions: For every infinitely divisible distribution, there is an associated Lévy process. Also, an infinitely divisible random variable can be thought of as a sum of a large number of independent and identically distributed shocks. These concepts will be unpacked more quantitatively in the subsequent chapters, but for now it should suffice to have a very basic grasp of what a Lévy process is and how these more generally extend beyond Brownian motions. The following properties of Lévy processes strengthen arguments for their use in the world of financial modelling [37]:

- They are consistent with the assumption of no arbitrage.
- They provide infinitely divisible distributions, which provides a means to express price changes as resulting from a great number of shocks to the economy.
- In order to have finite variation Lévy processes (which is possible), the diffusion component of a Lévy process must be zero and the process must be a pure jump Lévy process (in contrast to a jump-diffusion Lévy process)

Using a framework built on Lévy processes yields models that are more flexible than the standard geometric Brownian motion models. It is worth mentioning that with an increase in flexibility (degrees of freedom) of a model comes a tradeoff. The more – or better – one would like a model

to 'fit' observed data, the more flexible one has to make it. There is an inverse relationship between observed data and underlying structure of a model. Ultimately though, a feature of the pricing model used, is that it should have a 'backbone' or some feet to stand on, so that it can actually be of practical use. Striking the balance between naive models and useless models must be carefully considered. Even though a model may be perfectly able to match observed data, it may (through over-fitting for example) produce spurious and unreliable results – of no particular use to anyone.

There are a few different fundamental Lévy processes which can be (and are) used to model asset returns. All have their own unique properties and differences, but key to financial modelling is that they are able to back out prices for financial derivatives. Pricing generally does not happen through the same simple Black-Scholes closed form integral solution – other techniques are required. One approach [21] involves using an inverse Fourier transform on an analytic expression for the Fourier transform of the option or its time value – the expression is derived in [21]. More elaborate Monte Carlo techniques are also well suited to this problem and are widely used. Thus, a greater reliance on efficient numerical techniques is required and approximations for simulations and pricing become important. Fortunately, in many cases Lévy processes have tractable characteristic functions – a good starting point. There are two major types of Lévy processes: Jump-Diffusion processes and Infinite Activity processes. These results are summarised in Table 1.1. But the essential difference is philosophical: In Jump-Diffusion models the diffusion component accounts for high activity in the price process while the jump component is used to account for rare and extreme price movements. Infinite Activity processes, on the other hand, unify these two ideas: high activity is accounted for by an infinite number of small jumps, and the frequency of larger jumps is smaller than the frequency of small jumps - as a requirement. The expected number of jumps in a given interval per unit time is specified through a quantity called the *Lévy measure*. Although these models share the infinite amount of small movements, it is not necessarily the case that the total variation of the process on any bounded interval will be infinite, as it is with the Jump-Diffusion models. The variance-gamma process is an example of this – it is an infinite activity process with finite variation.

Recent empirical studies have speculated that diffusion components may not be strictly necessary in modelling asset prices [20]. One may be tempted to think that jumps arrive at discrete times and that in between those jump times it may suffice to model the process as a regular diffusion. This is known as jump-diffusion modelling, however these models are not widely used. One

Jump-Diffusion models	**Infinite Activity models**
Must contain a Brownian component	Do not necessarily contain a Brownian component
Jumps are rare events	Process basically moves by jumps
Distribution of jump sizes is known	Distribution of jump sizes does not exist - jumps arrive infinitely often
Perform well for implied volatility smile interpolation	Give realistic description of the historic price process
Densities are not known in closed form	Closed form densities sometimes known
Easy to simulate	Some have representation by Brownian subordination - additional tractability

TABLE 1.1: Summary of main differences between Jump-Diffusion models and Infinite Activity models [25]

reason for their unpopularity a choice of model is that no matter what ingredients or optional extras are thrown into a diffusion model (stochastic volatility, time and state dependent local volatility or something else) it will, with probability 1, generate (price) processes which are continuous functions of time. The presence of discontinuities in the observed price processes should be enough then to argue that these diffusion models will not be realistic representations [25].

Additionally, the very nature of volatility under the use of Lévy processes is that it is stochastic [25]. Volatility, a term which is flung around rather loosely these days, has many confusing connotations, and care has to be taken in defining it and deciding which measure of volatility should be used. Recently, many authors have turned towards use of a model-free notion of quadratic variation as a measure of market volatility - the so called "realised volatility":

$$v_\Delta(T) = \sum_{t=1}^{N} |r_t(\Delta)|^2 = \sum_{t=1}^{N} |X_{t+\Delta} - X_t|^2 \tag{1.1}$$

This differs from the sample variance of returns:

$$\hat{\sigma}^2(\Delta) = \frac{1}{N}\sum_{t=1}^{N} |r_t(\Delta)|^2 - \left[\frac{1}{N}\sum_{t=1}^{N} |r_t(\Delta)|\right]^2 \tag{1.2}$$

The serial uncorrelation of returns has the implication that the sample variance scales linearly with time: $\hat{\sigma}^2(\Delta) = \Delta\hat{\sigma}^2(1)$. The "realised volatility" can be shown ([25], Chapter 8) to converge (in probability) to a nontrivial stochastic process $([X,X]_t)_{t\in[0,T]}$ - the quadratic variation of X. If X is then a Lévy process with triplet (σ^2, ν, γ) then the quadratic variation process is given by:

$$[X]_t = \sigma^2 t + \sum_{0\le s\le t} |\Delta X_s|^2 \tag{1.3}$$

Thus $[X]_t$ is a random variable - it is only deterministic if there are no jumps (when log price is a Brownian motion with drift). So the assumption of a Lévy process which is not a Brownian motion means that realised volatility is *always* stochastic. One does not need to insert an additional parameter, or family of parameters, to allow for these effects.

Lévy processes offer insight into many of the features left untouched or unquestioned by the Black-Scholes-Merton framework. Although, mathematically, they are not as simple or easy to use, the additional efforts required to understand them may well pay off in the long run. This brings forth the issue of simulation, which forms the main crux of this dissertation. Efficient simulation should not be underestimated, it is key to getting it right – and that is the main focus of this dissertation, as discussed in the next section.

1.5 The issue of simulation and pricing

Pricing financial instruments is a subject which lies at the very core of financial mathematics. Although pricing and simulation may come across as separate issues, simulation often is a precursor to pricing, particularly when closed form solutions for pricing is not available. When valuing financial instruments a risk-neutral framework must be used, so as to avoid arbitrage issues. This is known as risk-neutral valuation and the following points are worth mentioning in this regard: Pricing is possible because of replication in complete markets and because of no-arbitrage bounds in incomplete markets. A model is said to be arbitrage-free if and only if at least one risk-neutral measure exists. (If this risk-neutral measure is unique, then the model

is said to be complete.) Since, under a risk-neutral measure, discounted prices are martingales, the prices of claims can be expressed as expectations. This makes it possible to price claims by means of a very general method known as *Monte Carlo* pricing. Its ability to price virtually any contingent claim structure makes it a very valuable and important tool in financial mathematics. Simulation from the underlying distribution of the asset is crucial to Monte Carlo pricing. Monte Carlo can even be used in the instance of the Black-Scholes model, where puts and calls can be priced by means of simulation. Certain other types of exotic options, such as asian options, lookback options and barrier options amongst others, require knowledge of the entire (or parts of) the asset path – this means that the entire asset path needs to be simulated. Another example of more recent claims which require simulation are options on realised volatility. Good sources of introductory reading on Monte Carlo techniques include Gentle [39], Jäckel [49] and Glasserman [40].

The implementation of Monte Carlo pricing can only be done once a particular process has been chosen, under appropriate risk-neutral dynamics. This is referred to as the *calibration problem*: each process is uniquely determined by a specific set of parameters, and these parameters require estimation. Examples of parameters requiring calibration are the volatility, skewness and kurtosis of a chosen process. Either a transformation from observed, real world dynamics to risk-neutral dynamics must be used, or a method must be used to extract risk-neutral dynamics out of current market information. Popular ways of calibrating models include backing out an implied risk neutral distribution from current market derivative prices, using the result of Breeden & Litzenberger [18]. Implementing it in markets such as the South African one requires careful thought however, since a large number of quoted strike prices are typically unavailable in this market. This may affect the implied distribution, which has to be backed out by means of rigorous interpolation schemes. The implied risk-neutral distribution then allows for estimation of the risk-neutral model parameters, which can be used to simulate increments of the risk-neutral process. A "minimal entropy least squares" is another approach considered in a Ph.D thesis by Peter Tankov [79]. Model calibration is a separate subject by itself, and is not considered further in this dissertation.

The risk-neutral measure obtained from calibration should not be unique as in the case of the Black-Scholes model. This is because markets under Lévy processes are not considered to be complete. Monte Carlo pricing takes over when sample paths of the risk-neutral distribution can be simulated.

1.5.1 A mathematical description of Monte Carlo methods

Monte Carlo integration techniques can be used to find the expected value of some function $f(x)$ with respect to some distribution density $\psi(x)$ over $x \in \mathbb{R}^n$. This can be stated mathematically as:

$$v = \mathbb{E}_{\psi(x)}[f(x)] = \int_{\mathbb{R}^n} f(x)\psi(x)dx^n \tag{1.4}$$

Assuming a method of drawing variates from the target distribution $\psi(x)$ exists, Monte Carlo integration is carried out as follows [40, 49]: calculate (or draw) a vector x_i, and use this to calculate $f_i = f(x_i)$, where f denotes the payoff function. By repeating this procedure, one can calculate, amongst other things, the running average which provides a Monte Carlo estimator:

$$\hat{v}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \tag{1.5a}$$

The above process is then repeated until either a predetermined number of iterations have been implemented or until a specific error estimate is reached. Since the Monte Carlo estimator is an average of many independent, individual draws of a random variate, say V, the central limit theorem ensures that, for large n, each individual evaluation of the estimator itself behaves as a normally distributed random variable,

$$\hat{v}_n \xrightarrow{i.d} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{1.5b}$$

if it is assumed that the variance of V is σ^2. Thus, one appropriate measure of uncertainty that can be used for the estimator $\hat{v}_n$ is the standard deviation of $\hat{v}_n$:

$$\sqrt{\text{Var}[\hat{v}_n]} = \frac{\sigma}{\sqrt{N}} \tag{1.5c}$$

In general, since the variance of V is typically unknown, it is often estimated by the variance of the simulation, due to the continuous mapping theorem[4] [49] along with the central limit theorem:

$$\hat{\sigma}_n = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} v_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)^2} \tag{1.5d}$$

This leads to a definition of the standard error:

$$\varepsilon_n = \frac{\hat{\sigma}_n}{\sqrt{n}} \tag{1.5e}$$

From this form of the standard error one identifies the relationship between the convergence of a Monte Carlo process and the number of simulations. Large simulation samples are not uncommon with the use of pseudo-random number generators, which converge as $\mathcal{O}(1/\sqrt{n})$ (as outlined above) implying that superior convergence requires additional simulations. It is possible to achieve better convergence rates using low discrepancy sequences (or Quasi Monte-Carlo (QMC) methods. Typically convergence rates for QMC simulations are much closer to n^{-1}, with the theoretical best being $\mathcal{O}(1/n)$ and the worst being $\mathcal{O}(c(d)[\ln(n)^d/n])$ [49]. The coefficient $c(d)$, however, depends on the dimensionality which makes estimation of this upper bound difficult. It is thus not straightforward to determine in advance whether QMC convergence will be quicker in higher dimensions. Other techniques aimed at improving convergence rates are stratified sampling, antithetic sampling and Latin Hypercube sampling. Adapting models to accommodate QMC methods for variance reduction is not as simple as in the Brownian case, and can easily lead to annoying and messy problems.

As can be seen, reducing the variance in Monte Carlo calculations is an important issue. Different claims generally require different variance reduction methods – for example different control variates must be used depending on the claim being priced. Thus, although pricing claims is the end goal, the concern of this thesis is not with the actual pricing, but rather with quick and efficient simulation of price paths.

Multi-dimensional financial problems require multi-dimensional integrals, and it is useful to give an indication as to what the typical dimensionality of a problem is equal to. A simple case considering the return of a portfolio given a multi-dimensional joint distribution for returns of the underlying assets for a specific time horizon would have dimensionality equal to the number

[4] Given a sequence (X_n, Y_n) which converges in distribution to (X, Y) and a continuous map ϕ, then $\phi(X_n, Y_n)$ converges in distribution to $\phi(X, Y)$

of underlying assets in the portfolio. When evaluating the expectation of a function(al) of discretised asset paths for a set of financial assets over a certain number of monitoring dates, then the dimensionality extends to $d = k \cdot l$ where k is the number of assets and l represents the number of time horizons (or monitoring dates).

There are two major books in the literature on Lévy processes: Schoutens [76] and Cont & Tankov [25] and the latter has been very valuable for this dissertation. On the subject of Monte Carlo finance the two books which have proved valuable are Jäckel [49] and Glasserman [40]. This dissertation focuses on providing a range of useful techniques for simulating from various Lévy processes and to approximate those Lévy processes which are difficult to simulate directly. Chapter 2 looks at some theoretical aspects and results pertaining to Lévy processes. Chapter 3 looks at some properties of specific Lévy processes for asset price modelling. Simulation from specific statistical distributions is the subject of Chapter 4 while Chapter 5 deals with the simulation of certain Lévy processes. Various results are then presented in Chapter 6.

Chapter 2

Lévy Processes

Before giving a formal definition of a Lévy process, it is worth defining the notion of a càdlàg function: In mathematics, a càdlàg function (continue à droite, limite à gauche) is a function, defined on the real numbers, which is everywhere right-continuous with left limits. A simple example of such a function would be a step function:

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \tag{2.1}$$

These functions reside in what is known as Skorokhod space. All continuous functions are also càdlàg and these kinds of functions play important roles in the theory of Lévy processes and appear throughout financial mathematics. The definition of a Lévy process now follows.

Definition 2.1. A Lévy process is a càdlàg process $(X_t)_{t\geq 0}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking on values in $\mathbb{R}^d$ and satisfying the following criteria [25]:

1. Independent Increments: For all increasing sequences of times $t_0, \ldots, t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.
2. Stationary Increments: the law of $X_{t+h} - X_t$ does not depend on t, i.e $X_{t+h} - X_t \sim X_h$.
3. Stochastic Continuity: $(\forall\, \varepsilon > 0)\ (\lim_{h\to 0} \mathbb{P}[|X_{t+h} - X_t| \geq \varepsilon] = 0)$.

The last condition does not imply that the sample paths are continuous. It is merely a formalization of the idea that for a given time t, the probability of seeing a jump is zero. In other words, the jumps are not predictable - they occur at random times. These exclude processes which have jumps at fixed times – such "calendar effects" are not relevant for the task at hand.

It is straightforward to verify that Brownian motion satisfies the above conditions, making it an example of a Lévy process. The class of Lévy processes is bigger than Brownian motions though and includes Poisson processes and Compound Poisson processes among many others. The *stationary increments* property implies that $X_0 = 0$ a.s.

The Lévy process has been introduced as an alternative model of asset returns, as mentioned in Chapter 1. This is because, empirically, the assumption that asset returns are lognormal is a weak assumption, and their price paths display marked discontinuities. In addition to this, volatility is not constant and this already has to be corrected for by a market skew. The Black-Scholes model is flawed, and so this dissertation investigates the sample paths of Lévy processes.

2.1 General properties of Lévy processes

2.1.1 Infinite Divisibility

THEOREM 2.1. Let $(X_t)_{t\geq 0}$ be a Lévy Process. Then for every t, X_t has an infinitely divisible distribution. Conversely, if F is an infinitely divisible distribution, then there exists a Levy Process (X_t) such that the distribution of X_1 is given by F.

Proof. The proof that $(X_t)_{t\geq 0}$ is a Lévy process $\Rightarrow$ X_t is infinitely divisible is given, for the converse statement the reader is referred to a proof in Cont & Tankov (2004) [25] (also see Sato (1999), Corollary 11.6 [74]).

Let $t > 0$ and $n \in \mathbb{N}$. Then X_t has representation

$$X_t = (X_{t/n} - X_0) + (X_{2t/n} - X_{t/n}) + \cdots + (X_t - X_{t(n-1)/n}) \tag{2.2}$$

which is a sum of independent identically distributed random variables. □

The importance of this result is straightforward. It implies a connection between Lévy processes and certain statistical distributions. These distributions are called *infinitely divisible distributions.*

Definition 2.2 (Infinite Divisibility)**.** A probability distribution F on $\mathbb{R}^d$ is said to be **infinitely divisible** if for *any* integer $n \geq 2$, there exist n independent and identically distributed (i.i.d) random variables $Y_1, \ldots, Y_n$, such that $Y_1 + \cdots + Y_n$ has distribution F.

Example 2.1. Consider the normal distribution: If Y_k are i.i.d. variables with $Y_k = X_{t/n} \sim N(\mu/n, \sigma^2/n)$ then $\sum_k Y_k \sim N(\mu, \sigma^2)$. This is a consequence of the addition of normally distributed random variables. Since n is arbitrary this relation holds for all $n > 0$ and the normal distribution is said to be *infinitely divisible.*

Note above that the normal distribution is closed under convolution. When two normally distributed random variables are added, another normally distributed random variable is obtained. It is not a requirement, however, that infinitely divisible distributions be closed under convolution. What is important is the realisation that infinitely divisible distributions can be split into arbitrarily many identically distributed 'bits'. The characteristic function of these arbitrarily many bits must then, mathematically, be the n^{th} convolution root of the infinitely divisible distribution, which can be determined though studying the characteristic functions of these infinitely divisible distributions. The infinite divisibility property can be shown to hold true for the following distributions as well: *gamma, Poisson, inverse Gaussian, lognormal, Pareto, Gumbel, Weibull, Student's-t, α-stable distributions.* This can be noted by looking at the characteristic functions of these distributions, which can be found in Sato [74]. A trivial example of a distribution which is not infinitely divisible is the uniform distribution.

2.1.2 Characteristic Functions

Proposition 2.1 (The Characteristic Function of a Lévy Process)**.** Let $(X_t)_{t\geq 0}$ be a Levy Process on $\mathbb{R}^d$. Then, there exists a continuous function $\psi : \mathbb{R}^d \to \mathbb{R}$ known as the *characteristic exponent* of X, such that:

$$\mathbb{E}[e^{iz\cdot X_t}] = e^{t\psi(z)} \tag{2.3}$$

Proof. Denote the characteristic function of $(X_t)_{t\geq 0}$ by,

$$\Phi_{X_t}(z) \equiv \mathbb{E}\left[e^{iz\cdot X_t}\right], \quad z \in \mathbb{R}^d$$

Writing X_{t+s} as $X_s + X_{t+s} - X_s$ decomposes X_{t+s} into two independent terms, implying the map from $t \to \Phi_{X_t}(z)$ is multiplicative:

$$\begin{aligned}\Phi_{X_t}(z) &= \Phi_{X_s}(z)\Phi_{X_{t+s}-X_s}(z) \\ &= \Phi_{X_s}(z)\Phi_{X_t}(z)\end{aligned}$$

Stochastic continuity of X_t implies $X_s \to X_t$ (in distribution) whenever $s \to t$. Because of this convergence in distribution it follows that $\Phi_{X_s}(z) \to \Phi_{X_t}(z)$ when $s \to t$ so $t \to \Phi_{X_t}(z)$ is continuous in t. Combining this with the multiplicative property implies that $t \to \Phi_{X_t}(z)$ is an exponential function [25, 35]. □

It is worth noting that the only freedom one has in specifying a Levy Process is to specify the law at one single, specific time. The law of X_t is determined by knowing what the law is at a time equal to 1. Furthermore, Lévy Processes are uniquely determined by their *characteristic triplets* and the *characteristic exponent* has a more specific representation which will be discussed later. A brief digression now follows, as the two most simple Lévy processes are defined. These processes are important because they are the building blocks of all other kinds of processes.

2.2 Poisson and Compound Poisson Processes

Definition 2.3 (Poisson Process). *Let $(\tau_i)_{i>1}$ be an increasing sequence of independent exponentially distributed random variables with parameter λ. Define $T_n = \sum_{i=1}^{n} \tau_i$. Now, define the process N_t by*

$$N_t = \sum_{n \geq 1} I_{\{T_n \leq t\}} \tag{2.4}$$

N_t is then called a Poisson Process with intensity λ.

The Poisson process itself may be expressed in terms of the random measure itself as follows:

$$N_t(\omega) = M(\omega, [0,t]) = \int_{[0,t]} M(\omega, ds) \tag{2.5}$$

The measure M also has an interpretation as the 'derivative' of a Poisson process – the Radon Nikodym derivative given by an increasing process. The derivative of a Poisson process is simply a superposition of Dirac masses located at the jump times, i.e

$$\frac{d}{dt} N_t(\omega) = M(\omega, [0,t]) = \sum_{i \geq 1} \delta_{T_i(\omega)} \tag{2.6}$$

As mentioned, the Poisson process is the simplest Lévy process, and is *not* a good choice for modelling asset returns. This is because (as will be shown below) the jumps, although occurring stochastically, have a deterministic amplitude of +1. Their simplistic structure, however, makes them a good starting point to clarify many of the theoretical concepts introduced in this Chapter. Additionally, their importance can be attributed to the fact that every Lévy process has a Compound Poisson process (CPP) representation when jumps larger than some $\varepsilon > 0$ are truncated.

The Poisson process starts at zero a.s. As t increases, the value of N_t will remain zero until the time value $t = \tau_1$ is reached. At this point the indicator function for that particular set $\{T_1 \leq t\}$ takes on the value of 1 so that $N_t = 1$. Notice that the sum is over n, for each t. For $t > T_1$ the value of N_t is still 1, up until the value of t for which $t = T_2$. For this particular value of

t, $t \in \{T_1 \leq t\}$ and $t \in \{T_2 \leq t\}$, so the value of the Poisson Process is now $N_t = 1 + 1 = 2$, since there are now two sets which t belongs to. The Poisson Process can thus intuitively be seen to be a counting process: It counts the number of sets to which every value of t belongs. Due to the nature of the construction of $(\tau_i)_{i>1}$ (it is a sum of positive variables) the sequence is increasing and so the process N_t is also increasing, with its value changing by $+1$ each time. Even though jump sizes are predictable, they occur at unpredictable times and hence the process is stochastic. Compound Poisson processes generalise this concept by sampling jumps from more general statistical distributions.

Proposition 2.2. *(Properties of Poisson Processes)*:
Let (N_t) be a Poisson Process with intensity λ, then:

1. For all $t > 0$, N_t is a.s finite.
2. For all ω the sample paths $t \to N_t$ are piecewise constant and increase by jumps of size $+1$.
3. Sample paths are càdlàg implying that $N_{t+} = N_t$.
4. For any $t > 0$ $N_{t-} = N_{t+}$ with probability 1, thus $N_{t-} = N_t$ a.s.
5. N_t is continuous in probability: $N_s \xrightarrow{\mathbb{P}} N_t$ as $s \to t$ for all $t > 0$
6. For any $t > 0$, N_t is Poisson distributed with parameter λt:
   i.e $\mathbb{P}(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$ for all $n \in \mathbb{N}$
7. N_t has independent increments: For any $t_1 < \cdots < t_n$, $N_{t_n} - N_{t_{n-1}}, \ldots, N_{t_2} - N_{t_1}, N_{t_1}$ are independent random variables.
8. The increments are homogeneous: for any $t > s$, $N_t - N_s$ has the same distribution as N_{t-s}.
9. (N_t) has the Markov Property.

Proofs of these claims are short and straightforward and can be found in Cont & Tankov [25].

Definition 2.4 (Characteristic Function of a Poisson Process)**.** Let $(N_t)_{t\geq 0}$ be a Poisson process with intensity parameter λ. Its characteristic function is given by:

$$\mathbb{E}[e^{iuN_t}] = \exp\{\lambda t(e^{iu} - 1)\} \tag{2.7}$$

This definition follows directly from the fact that $N_t \sim \text{Poisson}(\lambda t)$.

2.2.1 Forming new Poisson Processes from old ones

- **Sum of independent Poisson processes:**
  If $(N_t^1)_{t\geq 0}$ is a Poisson Process with intensity λ_1, and $(N_t^2)_{t\geq 0}$ is a Poisson Process with intensity λ_2 then $(N_t^1 + N_t^2)_{t\geq 0}$ is also a Poisson Process with intensity $\lambda_1 + \lambda_2$. This can easily be seen by looking at the characteristic function for $M_t = N_t^1 + N_t^2$:

$$\begin{aligned}
\mathbb{E}[e^{iuM_t}] &= \mathbb{E}[e^{iuN_t^1 + iuN_t^2}] \\
&= \mathbb{E}[e^{iuN_t^1}]\mathbb{E}[e^{iuN_t^2}] \\
&= \exp\{\lambda_1 t(e^{iu} - 1)\}\exp\{\lambda_2 t(e^{iu} - 1)\} \\
&= \exp\{(\lambda_1 + \lambda_2)t(e^{iu} - 1)\}
\end{aligned}$$

- **'Thinning' Poisson processes:**
  If (N_t) is a Poisson Process which has intensity parameter λ, and (X_t) is defined by taking all the jumps of (N_t) and keeping them with probability $p \in (0,1)$ (or discarding them with probability $1-p$). By ordering the retained jumps and defining:

$$X_t = \sum_{n \geq 1} I_{\{\tau'_n \leq t\}} \tag{2.8}$$

  X_t is now a Poisson Process with parameter $p\lambda$.

- **'Compensating' the small jumps:** If $(X_t)_{t \geq 0}$ is a Lévy process with $\mathbb{E}[|X_t|] < \infty$ then $X_t - \mathbb{E}[|X_t|]$ is a martingale [25], motivating the definition of a new Poisson process below.

Definition 2.5 (Compensated Poisson Process). *Let $(X_t)_{t \geq 0}$ be a Poisson process (implying $\mathbb{E}[|X_t|] < \infty$). Then:*

$$\tilde{N}_t = N_t - \lambda t \tag{2.9}$$

defines a Compensated Poisson process.

The characteristic function, through a straightforward calculation, looks as follows:

$$\Phi_{\tilde{N}_t} = \exp[\lambda t(e^{iz} - 1 - iz)] \tag{2.10}$$

The compensated Poisson process has independent increments and the quantity λt, which is deterministic, is known as the compensator. Note that the process is no longer integer valued and therefore not a counting process.

It can also be shown that:

$$\mathbb{E}\left[\frac{\tilde{N}_t}{\lambda}\right] = 0 \qquad\qquad Var\left[\frac{\tilde{N}_t}{\lambda}\right] = t \tag{2.11}$$

and so when the intensity of its jumps increases an (interpolated) compensated poisson process converges *in distribution* to a Wiener process – this follows from a 'functional' version of the central limit theorem on the Skhorohod[1] space ($\Omega = D([0,T])$ which is known as the *Donsker invariance principle* [25].

$$\left(\frac{\tilde{N}_t}{\lambda}\right)_{t \in [0,T]} \longrightarrow (W_t)_{t \in [0,T]} \tag{2.12}$$

2.2.2 Compound Poisson Process

It is possible to generalise Poisson processes introduced in the previous section so that they may be used for realistic asset price modelling. This is done through the idea of a *compound Poisson process*, which generalises a Poisson process by allowing jumps to come from any arbitrary distribution.

Definition 2.6 (Compound Poisson Process). : *Let $(X_t)_{t \geq 0}$ be a càdlàg stochastic process defined by:*

$$X_t = \sum_{i=1}^{N_t} Y_i \tag{2.13}$$

where (N_t) is a Poisson process with intensity λ independent of (Y_i), the jump size process, whose jumps, Y_i, are i.i.d with distribution f. $(X_t)_{t \geq 0}$ is said to be a Compound Poisson process with intensity λ.

[1] Generally, this is the space of càdlàg functions from $[0, \infty)$ to a metric space (E, d)

The distribution used for jump sizes is predetermined. Jump sizes (when they occur) are then sampled as *i.i.d* variates from this distribution. This allows for modelling flexibility and becomes important when compound Poisson processes are used as approximations to Lévy processes.

Proposition 2.3 (Characteristic function of a Compound Poisson process)**.** *Let $(X_t)_{t\geq 0}$ be a Compound Poisson process on $\mathbb{R}^d$. Its characteristic function has the following representation:*

$$\mathbb{E}[\exp(iu \cdot X_t)] = \exp\left\{t\lambda \int_{\mathbb{R}^d} (e^{iu\cdot x} - 1) f(dx)\right\}, \quad (\forall u \in \mathbb{R}^d) \tag{2.14}$$

which, in a one dimensional case can be written as

$$\mathbb{E}[\exp(iuX_t)] = \exp\left\{t \int_{-\infty}^{\infty} (e^{iu\cdot x} - 1) \nu(dx)\right\}, \quad (\forall u \in \mathbb{R}) \tag{2.15}$$

where $\nu = f\lambda$ is the Lévy measure of the process $(X_t)_{t\geq 0}$ [25].

Proof. Start by conditioning the expectation on N_t. Denote the characteristic function of f by $\hat{f}$:

$$\mathbb{E}[\exp\{iuX_t\}] = \mathbb{E}[\mathbb{E}[\exp\{iuX_t\}]|N_t] = \mathbb{E}[(\hat{f}(u))^{N_t}] \tag{2.16}$$

$$= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n (\hat{f}(u))^n}{n!} = \exp\{\lambda t(\hat{f}(u) - 1)\} \tag{2.17}$$

$$= \exp\left\{t\lambda \int_{\mathbb{R}^d} (e^{iu.x} - 1) f(dx)\right\} \tag{2.18}$$

□

2.3 Jumps and discontinuities

Sample paths of Lévy processes are not generally continuous. Knowing the *Lévy measure* of a process is equivalent to a complete understanding of the jump structure. This section discusses the various types of measures that are associated with Lévy processes, and these are used later in the Lévy -Ito decomposition theorem – as well as the Lévy -Khinchin representation.

2.3.1 Random Measures

Definition 2.7 (Random Measure)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(E, \mathcal{B})$ be a measurable space. A map $M : \mathcal{B} \times \Omega \mapsto \mathbb{R}$ is called a *random measure* on $(E, \mathcal{B})$ iff

1. For each $B \in \mathcal{B}$, the map $\omega \mapsto M(B, \omega)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.
2. For almost every $\omega \in \Omega$, the map $B \mapsto M(B, \omega)$ is a measure on $(E, \mathcal{B})$.
3. There exists a partition of E, $B_1, B_2, \cdots \in \mathcal{B}$ such that $M(B_k) < \infty$ almost surely for all k.

Random measures are said to have independent increments iff $M(B_1), \ldots, M(B_n)$ are independent random variables. A *point process* is a random measure on $(E, \mathcal{B})$ which takes on positive integer values (including ∞), i.e iff M is $\bar{\mathbb{Z}}^+$–valued. A *Poisson random measure* with intensity λ is a point process M, with independent increments such that for every $B \in \mathcal{B}$, $M(B)$ is a Poisson random variable with mean $\lambda(B)$. Here λ is a measure on $(E, \mathcal{B})$. In other words

$$\mathbb{P}(M(B) = k) = e^{-\lambda(B)} \frac{\lambda(B)^k}{k!} \tag{2.19}$$

To every càdlàg process a random measure, known as the jump measure can be assigned. Denote $A = (0, \infty) \times \mathbb{R}^d \backslash \{0\}$. If the process is a Lévy process, the jump measure is a Poisson random measure on $(A, \mathcal{B}(A))$ [25]. The jump measure is defined as follows:

Definition 2.8 (Jump Measure)**.** Let $(X_t)_{t \geq 0}$ be a càdlàg stochastic process on $(\Omega, \mathcal{F}_t, \mathbb{P})$ and $H \in \mathcal{B}(A)$. For every $\omega \in \Omega$, define J_X the jump measure of the process X_t in the following way:

$$J_X(\omega, .) = \#\{t : (t, \Delta X_t) \in H\} = \sum_{t \in [0,T]}^{\Delta X_t \neq 0} \delta_{(t, \Delta X_t)} \tag{2.20}$$

Intuitively speaking, for any set $H \subset \mathbb{R}^d$, the jump measure $J_X([0, t], H)$ counts the number of jumps between 0 and t, whose amplitude is a member of the set H. These measures are used in the Lévy -Ito decomposition theorem and therefore have a role to play in approximating all kinds of Lévy processes.

2.3.2 The Lévy measure

One of the defining features of a Lévy process is its Lévy measure. This measure is defined to be a positive measure on $\mathbb{R}^d$, which need not necessarily be a probability measure. It describes the jump structure of a Lévy process entirely. If X_t is a Lévy process, then because it is càdlàg it is possible to define $\Delta X_t = X_t - X_{t-}$. It is quite possible for the sum $\sum_{s \leq t} \Delta X_s$ to be infinite, but for any bounded time interval there can be only finitely many jumps whose amplitude exceed a certain size – again because X_t is càdlàg [25]. A set $B \in \mathcal{B}$ is said to be bounded away from zero if $0 \notin \bar{B}$. For any such B, the sum

$$\sum_{s \leq t} \Delta X_s \mathbb{I}_{\Delta X_s \in B}$$

will have only finitely many non zero terms and a strictly increasing sequence of stopping times $(\tau_n^B)_{n \in \mathbb{N}}$ can be introduced as follows:

$$\tau_0^B = 0 \qquad \tau_{n+1}^B = \inf\{t > \tau_n^B : \Delta X_t \in B\}$$

These then enumerate the jump times in B. Let $N_t(B)$ be the associated counting process

$$N_t(B) = \sum_{n=1}^{\infty} \mathbb{I}_{\{\tau_n^B \leq t\}} = \sum_{0 \leq s \leq t} \mathbb{I}_B(\Delta X_s)$$

Now introduce the parameter of $N_t(B)$, $\nu(B)$ as

$$\nu(B) = \mathbb{E}[N_1(B)] \tag{2.21}$$

This is simply the expected number of jumps of $N_t(B)$ per unit time. Note then that $\nu(B)$ also gives the expected number of jumps of X which belong to B per unit time. It is straightforward to verify that ν satisfies the conditions of a measure on $\mathcal{B}(\mathbb{R}\backslash\{0\})$. Also from the definition of ν in (2.21) above, the monotone convergence theorem implies that ν is also a measure.

Definition 2.9 (Lévy measure)**.** Let $(X_t)_{t \geq 0}$ be a Lévy process on $\mathbb{R}^d$. The measure ν defined on $\mathbb{R}^d$ defined by:

$$\nu(A) = \mathbb{E}[\#\{t \in [0, 1] : \Delta X_t \neq 0, \Delta X_t \in A\}], \quad A \in \mathcal{B}(\mathbb{R}^d) \tag{2.22}$$

is known as the Lévy measure of X. Above, # represents the *counting measure*, counting the number of elements in a given set. $\nu(A)$ represents the expected number of jumps with size belonging to A, per unit time.

2.4 The Lévy -Ito decomposition theorem

THEOREM 2.2. *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ and ν its Lévy measure, given by Definition (2.9). Then,*

- *ν is a positive measure on $\mathbb{R}^d$ satisfying*

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty \qquad \int_{|x|>1} \nu(dx) < \infty \tag{2.23}$$

- *The jump measure of (X_t), J_X is a Poisson random measure on $[0,\infty)\times\mathbb{R}^d$ with intensity measure $\nu(dx)dt$*

- *there exists a constant vector γ and a d-dimensional Brownian motion with (constant) covariance matrix A, such that*

$$X_t = \gamma t + B_t + X_t^l + \lim_{\varepsilon\downarrow 0} \tilde{X}_t^\varepsilon \tag{2.24}$$

with,

$$\begin{aligned} X_t^l &= \int_{|x|\geq 1, s\in[0,t]} x J_X(ds\times dx) \\ \tilde{X}_t^\varepsilon &= \int_{\varepsilon\leq|x|<1, s\in[0,t]} x\{J_X(ds\times dx) - \nu(dx)ds\} \\ &\equiv \int_{\varepsilon\leq|x|<1, s\in[0,t]} x\tilde{J}_X(ds\times dx) \end{aligned}$$

A rigorous proof can be found in Sato [74] with an outline given by Cont & Tankov [25]. The proof, initially found by Lévy , was completed by Ito [25]. For purposes of this dissertation the above theorem is simply listed. The first two terms in (2.24) are none other than a Brownian motion with a drift term – these are the continuous terms. The last two terms represent the discontinuous (or jump) components of the Lévy process. The condition $\int_{|x|\geq 1}\nu(dx) < \infty$ implies that the number of jumps (of (X_t)) with absolute value greater than 1 is finite, so the term X_t^l is actually an *a.s* finite sum. Thus, X_t^l is a Compound Poisson process.

The value 1 could be replaced with any $\varepsilon > 0$, provided it's not 'close' enough to zero (in the neighbourhood of zero) and the resulting process X_t^ε would still be Compound Poisson. The problem arises for values of ε which are in the neighbourhood of zero since contrary to the Compound Poisson case, ν can have a singularity at zero. Infinite activity Lévy processes are an example of this, and so it is possible to have infinitely many small jumps, the sum of which does not converge. In order to obtain convergence in the limit as $\varepsilon\downarrow 0$, the remainder term X_t^ε must be replaced by its *compensated* version, where the expected value of the small jumps is subtracted off thus achieving convergence.

In equation (2.24) above, not all the terms are martingales; only the B_t term and the compensated jump term, $\tilde{X}_t^\varepsilon$. An important implication of the Lévy -Ito decomposition theorem is that every Lévy process can be approximated with arbitrary precision by a jump-diffusion process: A sum of a Brownian motion with drift and a Compound Poisson process – possibly more than one. This is a point of practical relevance, particularly as far as the simulation of Lévy processes is concerned.

2.5 The Lévy -Khinchin representation

Theorem 2.3 below gives a useful representation for the form of a characteristic function of a Lévy process. It can be simplified slightly in the case where the Lévy process is of finite variation, and gives insight into the parameters of a Lévy process. Although, it is not required for the purposes of this dissertation, it is worth stating because of its importance and its use in pricing techniques involving Fourier transforms. These methods are discussed in papers such as [21].

THEOREM 2.3. Let $(X_t)_{t\geq 0}$ be a real-valued Lévy process on $\mathbb{R}^d$. The characteristic function has the following representation:

$$\mathbb{E}[e^{iz\cdot X_t}] = e^{t\psi(z)}, \quad z \in \mathbb{R}^d \tag{2.25}$$

with

$$\psi(z) = -\frac{1}{2}z\cdot Az + i\gamma\cdot z + \int_{\mathbb{R}^d}(e^{iz\cdot x} - 1 - iz\cdot x\mathbb{I}_{|x|\leq 1})\nu(dx) \tag{2.26}$$

where A is a $d\times d$ matrix with positive, real valued entries. $\gamma \in \mathbb{R}^d$ and ν being a positive measure satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}}(1\wedge|x|^2)\nu(dx) < \infty$. The triplet (A,γ,ν) is known in literature as the *characteristic triplet*. The measure ν is the Lévy measure and plays an important role in the theory of Lévy processes.

Proof. Theorem 8.1 in Sato [74]. □

Jumps larger than some arbitrary $\varepsilon > 0$ may be truncated, as opposed to the above equation where they were truncated at 1. Generally speaking, the concept of a truncation function, g, can be introduced. This function must be a bounded, measurable function $g : \mathbb{R}^d \to \mathbb{R}$ satisfying $g(x) = 1 + o(|x|)$ as $x \to 0$ and $g(x) = O(1/|x|)$ as $x \to \infty$. The representation above then becomes

$$\psi(z) = -\frac{1}{2}z\cdot Az + i\gamma^g\cdot z + \int_{\mathbb{R}^d}(e^{iz\cdot x} - 1 - iz\cdot xg(x))\nu(dx) \tag{2.27}$$

$$\gamma^g = \gamma + \int_{\mathbb{R}^d} x(g(x) - \mathbb{I}_{|x|\leq 1})\nu(dx) \tag{2.28}$$

The choice of truncation function does not affect the values of A and ν (which are intrinsic parameters of the Lévy process), but it does affect γ. Thus, one should avoid calling γ the "drift" of the process.

2.5.1 Finite variation Lévy processes

The concept of variation of a function has a part to play when it comes to the classification of Lévy processes. The *total variation* of a function $f : [a,b] \to \mathbb{R}^d$ is defined as:

$$TV(f) = \sup\sum_{i=1}^{n}|f(t_i) - f(t_{i-1})| \tag{2.29}$$

where the supremum is taken over all finite partitions of the interval $[a,b]$. It turns out, [25], that a Lévy process is of finite variation iff its characteristic triplet (A,ν,γ) satisfies: $A = \mathbf{0}$ and $\int_{|x|\leq 1}|x|\nu(dx) < \infty$. This simplifies the Lévy -Khinchin representation as well as the Lévy -Ito decomposition.

2.5.2 Representation in the finite variation case

If $(X_t)_{t\geq 0}$ is a Lévy process of finite variation with Lévy -triplet $(0, \nu, \gamma)$ then for the Lévy -Ito decomposition we have [25]:

$$X_t = bt + \int_{[0,t]\times\mathbb{R}^d} xJ_X(ds \times dx) = bt + \sum_{s\in[0,t]}^{\Delta X_s \neq 0} \Delta X_s \tag{2.30}$$

While the characteristic function is given by:

$$\mathbb{E}[e^{iz\cdot X_t}] = \exp\left[t\left\{ib\cdot z + \int_{\mathbb{R}^d}(e^{iz\cdot x} - 1)\nu(dx)\right\}\right] \tag{2.31}$$

2.6 Subordinators

Definition 2.10 (Subordinator)**.** A Lévy process $(S_t)_{t\geq 0} \in \mathbb{R}$ is said to be a subordinator iff it satisfies one of the following equivalent (see [25]) properties.

1. $S_t \geq 0$ for some $t > 0$ (almost surely)
2. $S_t \geq 0$ for all $t > 0$ (almost surely)
3. Sample paths of S_t are almost surely non-decreasing: $t \geq s \Rightarrow S_t \geq S_s$
4. The characteristic triplet satisfies $A = 0$, $\nu((-\infty, 0]) = 0$, $\int_0^\infty (x \wedge 1)\nu(dx) < \infty$ and $b > 0$. In other words, S_t has no diffusion component, has positive jumps (of finite variation) and has positive 'drift'.

Put very simply, a subordinator, $(S_t)_{t\geq 0}$, is a non decreasing Lévy process. Since S_t is a positive random variable, it is possible (and more convenient) to work with the *Laplace transform*, as opposed to the Fourier transform. The moment generating function of $(S_t)_{t\geq 0}$ can then be written in terms of the *Laplace exponent*, $\ell(u)$, in place of the characteristic exponent , $\psi(u)$ [25]:

$$\mathbb{E}[e^{uS_t}] = e^{t\ell(u)} \quad (\forall u \leq 0), \qquad \ell(u) = bu + \int_0^\infty (e^{ux} - 1)\nu(dx) \tag{2.32}$$

Their characteristic triplets are then given by $(0, b, \nu)$. The next theorem illustrates their importance and influence in the theory of Lévy processes and time-changing.

THEOREM 2.4 (Subordination)**.** *Fix $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic exponent $\psi(u)$ and triplet (A, ν, γ). Let $(S_t)_{t\geq 0}$ be a subordinator with Laplace exponent $\ell(u)$ and triplet $(0, \rho, b)$. Then the process defined by:*

$$Y(t;\omega) = X(S(t;\omega);\omega) \quad \forall\omega \in \Omega \tag{2.33}$$

is again a Lévy process, with characteristic function given by:

$$\mathbb{E}[e^{iuY_t}] = \exp(t\ell(\psi(u))) \tag{2.34}$$

That is to say, the characteristic exponent of Y is obtained by composition of the Laplace exponent of S with the characteristic exponent of X. The characteristic triplet (A^Y, ν^Y, γ^Y) is given

by (p_t^X is the probability density function of X_t):

$$A^Y = bA \tag{2.35a}$$

$$\nu^Y(B) = b\nu(B) + \int_0^\infty p_s^X(B)\rho(ds) \quad \forall B \in \mathcal{B}(\mathbb{R}^d) \tag{2.35b}$$

$$\gamma^Y = b\gamma + \int_0^\infty \rho(ds) \int_{|x|\leq 1} x p_s^X(dx) \tag{2.35c}$$

The process $(Y_t)_{t\geq 0}$ is said to be subordinate to the process $(X_t)_{t\geq 0}$.

A very detailed proof can be found in Sato (1999), and for further reading a relatively casual discussion is given in Cont & Tankov (2004). The following theorem takes the groundwork laid above and applies it to the context of *Brownian subordination.* The result is a new Lévy process (constructed from an old one) with the new characteristic triplet given in terms of the old one.

THEOREM 2.5 (Characteristic Function of Lévy processes based on Brownian subordination)**.** *Let $X(t;\omega) = \theta t + \sigma B_t(\omega)$ be a Brownian motion with drift (a Lévy process on $\mathbb{R}^d$) with characteristic triplet $(\Sigma, 0, \theta)$. Let $S_t(\omega)$ be a subordinator independent of X_t on $\mathbb{R}$ with Laplace exponent $\ell(u)$ and characteristic triplet $(0, \nu, b)$.*

The process given by $Y_t(\omega) = X(S_t(\omega);\omega) = \theta S_t(\omega) + \sigma B(S_t(\omega);\omega)$ is a Lévy process with characteristic triplet given by $(\Sigma^S, \nu^S, \theta^S)$ where:

$$\Sigma^Y = b\Sigma \tag{2.36a}$$

$$\nu^Y(x) = \int_0^\infty f(x,s)\nu(s)ds \tag{2.36b}$$

$$\theta^Y = b\theta + \int_0^\infty \int_{|x|\leq 1} x f(x,s)\nu(s)dxds \tag{2.36c}$$

where $f(x,s)$ has a multivariate normal density with mean vector θs and covariance matrix $s\Sigma$. Also, the characteristic function of the process Y_t, is given by:

$$\mathbb{E}\left[e^{iuY_t}\right] = e^{t\ell(iu\theta - u^2\sigma^2/2)} \tag{2.37}$$

Proof. This follows as a result of Theorem 2.4. □

The fact that $f(x,s)$ has multivariate normal density is an important consequence of the fact that the subordination here is Brownian. These subordinators play a specific role in construction of Lévy processes from Brownian motions (with drift μ and variance σ^2) through a change from conventional to *stochastic* time. This time change affects the normality assumption of log returns.

Cumulants of a distribution are defined from its *cumulant generating function*: if a distribution has a characteristic function Φ_X, which satisfies $\Phi_X(0) = 0$ [25]. Because Φ_X is continuous at 0, $\Phi_X(z) \neq 0$ in a neighbourhood of 0, and so a continuous version of the logarithm of Φ can be defined in this region. This function Ψ_X satisfies

$$\Psi_X(0) = 0, \quad \Phi_X(z) = \exp[\Psi_X(z)]$$

Cumulants are then given by

$$c_n(X) = \frac{1}{i^n}\frac{\partial^n \Psi_X}{\partial u^n}(0) \tag{2.38}$$

The $n-$th cumulant can be expressed as a polynomial function of the moments m_k, $k = 1 \dots n$. The characteristic exponent of X, $\Psi(u) = \ell(-u^2\sigma^2/2 + i\theta u)$ allows cumulants of X to be calculated from those of S. In the symmetric case ($\theta = 0$) the mean and skewness of X_t are zero (since it is symmetric) and the variance and excess kurtosis are given by:

$$\mathrm{Var}[X_t] = \sigma^2 \mathbb{E}[S_t] \tag{2.39a}$$

$$\kappa[X_t] = \frac{3\mathrm{Var}[S_t]}{\mathbb{E}[S_t]^2} \tag{2.39b}$$

Therefore, X_t will be leptokurtic if S_t is not deterministic. Furthermore, Brownian subordination imposes certain limitations on the form associated with the Lévy measures. The following proposition characterises Lévy measures of processes allowing representation as Brownian motions with drift.

Proposition 2.4. *Let ν be a Lévy measure on $\mathbb{R}$ and $\mu \in \mathbb{R}$. There exists a Lévy process $(X_t)_{t\geq 0}$ with Lévy measure ν such that $X_t = W(Z_t) + \mu Z_t$ for some subordinator $(Z_t)_{t\geq 0}$ and some Brownian motion $(W_t)_{t\geq 0}$ independent from Z iff the following conditions are satisfied:*

1. *ν is absolutely continuous with density $\nu(x)$.*
2. *$\nu(x)e^{-\mu x} = \nu(-x)e^{\mu x}$ for all x.*
3. *$\nu(\sqrt{(}u))e^{-\mu\sqrt{u}}$ is a completely monotonic function on $(0, \infty)$.*

It is maintained that the small jumps of such processes are always symmetric, because the Lévy measure has a tilted version symmetric on $\mathbb{R}$. Exponential tilting mainly affects the big jumps. Furthermore, if ν is a Lévy measure on $\mathbb{R}$, then it can only be the Lévy measure of a subordinated Brownian motion (without drift) iff it is symmetric and $\nu(\sqrt{u})$ is completely monotonic on $(0, \infty)$.

2.7 Constructing Lévy processes

In order to use Lévy processes, a means of constructing them is required. The ensuing discussion aims to introduce some of the different available options, and show when their use is best suited and how these methods can be carried out. The starting point is to note that certain types of transformations exist which, when applied to a Lévy process, $(X_t)_{t\geq 0}$, produce another Lévy process given by $\mathcal{T}[(X_t)_{t\geq 0}] = (X'_t)_{t\geq 0}$. This transformation should affect the underlying parameters of the process in deterministic ways, meaning that given the initial parameters of the Lévy process it is possible to determine the new parameters under the transformation. These alterations are often of interest. Girsanov's theorem, a useful result stating that a Brownian motion under a change of measure is still a Brownian motion does not hold true, in general, for Lévy processes, and this is an important point to bear in mind when modelling.

Construction via Brownian subordination

If a Brownian motion is subordinated by another suitable Lévy process, Theorem 2.4 details how the parameters of the subordinated Brownian motion can be determined. Subordination can be defined quite generally, leading to a generalisation of the results in Theorem 2.4. In this dissertation, only Brownian subordination is considered, as it is of practical relevance to the problem at hand. The general procedure involves taking a subordinator, $(S_t)_{t\geq 0}$ (with Laplace exponent $\ell(u)$), an independent Brownian motion $(W_t)_{t\geq 0}$ and replacing the time variable in the Brownian motion with the stochastic process (S_t): $X_{S_t} = \mu S_t + \sigma W(S_t)$. Theorem 2.4

guarantees that a new Lévy process is obtained with interpretation that the time scale has been changed to *business time.*

The attraction here is that modelling of the arrival of information ('events') is inherently stochastic and not the simple, constant and familiar continuous flow of a Brownian motion. That the starting point is a Brownian motion promotes the significance of this approach for two reasons: firstly, it helps maintain the practical relevance of such Lévy process models, supporting their use above the class of general (and esoteric) Lévy processes, while still incorporating the 'random walk' aspect of financial modelling which has been present since the 1900's. The second reason is that these models have additional tractability and flexibility and are thus worth investigating as alternatives to the standard GBM models.

Use of a subordination approach allows an immediate and straightforward computation of the characteristic function . This is a consequence of the conditional Gaussian structure of these processes which also simplifies computations and simulations significantly. As an example, the call option price can be expressed as an integral involving Black-Scholes prices. However, an explicit form of the Lévy measure is not always available. These processes are worthwhile if one has a valid subordinating Lévy process and popular models such as the variance-gamma and normal inverse Gaussian are built on this framework.

Linear Transformations

Another way of transforming Lévy processes is to perform a standard linear transformation. The resulting changes that happen are given by the next theorem.

THEOREM 2.6. Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic triplet (A, ν, γ) and let M be an $n \times d$ matrix. Then $Y_t = MX_t$ is a Lévy process on $\mathbb{R}^n$ with characteristic triplet (A_Y, ν_Y, γ_Y), where:

$$A_Y = MAM^{tr} \tag{2.40}$$

$$\nu_Y = \nu(\{x : Mx \in B\}) \quad \forall B \in \mathcal{B}(\mathbb{R}^n) \tag{2.41}$$

$$\gamma_Y = M\gamma + \int_{\mathbb{R}^n} y(\mathbb{I}_{\{|y|\leq 1\}}(y) - \mathbb{I}_{S_1}(y))\nu_Y\, dy \tag{2.42}$$

S_1 is the image of the unit ball in $\mathbb{R}^d$ under transformation M

Example 2.2. Suppose $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are two independent Lévy processes with characteristic triplets given by (A_X, ν_X, γ_X), (A_Y, ν_Y, γ_Y). $X_t + Y_t$ is a Lévy process with

$$A = A_X + A_Y \tag{2.43}$$

$$\nu(B) = \nu_X(B) + \nu_Y(B) \quad (\forall B \in \mathcal{B}(\mathbb{R})) \tag{2.44}$$

$$\gamma = \gamma_X + \gamma_Y - \int_{[-\sqrt{2},-1]\cup[1,\sqrt{2}]} y\nu(dy) \tag{2.45}$$

[Use Theorem 2.6 with $M^{tr} = (1 \quad 1)$]

Exponential Tilting of the Lévy measure

One possible way of specifying a Lévy process is through specification of an admissible characteristic triplet. It is a requirement that the Lévy measure satisfies the following constraints:

$$\int_{|x|\leq 1} |x|^2\nu(dx) < \infty \qquad \int_{|x|\geq 1} \nu(dx) < \infty \tag{2.46}$$

Provided a transformation to the Lévy measure can be found respecting the above integrability conditions, a new Lévy process can be constructed. An example of such a transformation would be to multiply the Lévy measure by an exponential function. More formally: If there exists a $\theta \in \mathbb{R}^d$ such that $\int_{|x|\geq 1} e^{\theta \cdot x} \nu(dx) < \infty$, then

$$\tilde{\nu}(dx) := e^{\theta \cdot x} \nu(dx) \tag{2.47}$$

defines a new Lévy measure. This type of transformation is known as an *Esscher transform* and is also referred to as *exponential tilting* of the Lévy measure. The Esscher transform is somewhat analogous to the drift change transformation of geometric Brownian motion which means it can be used to find equivalent martingale measures in a Lévy model framework. If the Gaussian component is absent, the drift cannot be changed, but the jump distribution can still be altered leading, again, to a wide variety of equivalent measures [25].

Specification of the Lévy measure

This is done, for example, in the case of tempered stable processes. An advantage of this method is that one has an idea as to the pathwise nature of the process beforehand, through specification of the jump structure of the process. The Lévy -Khinchin formula allows determination of the distribution of the process at any time. Simulation is generally quite an involved process, but the use of this approach exposes the modeler to a wide variety of models.

Explicit specification of the probability density function for $t = \Delta$

This approach, is used for the general hyperbolic (GH) models – a broad class of models studied extensively by Barndorff-Nielsen ([8, 9, 10, 11] amongst others). It allows for easy simulation, provided that it is on a fixed time grid. Estimating parameters of the distribution is also simple if data sampling has been carried out at the same frequency. The Lévy measure however, is unknown, making knowledge of the law of increments at other time scales unknown. The nature of the process is also not immediately obvious (infinite or finite variation, infinite or finite activity, Gaussian component or not and so on).

Chapter 3

Lévy processes for asset modelling

A synopsis of certain different types of Lévy processes is now presented. Unlike Brownian motion, modelling with Lévy processes is both more flexible and versatile, because there are many different types of Lévy processes. Therefore, it is a good idea to have in mind the advantages and disadvantages associated with the different Lévy processes. In the ensuing chapter, the more common models such as the variance-gamma model and the normal-inverse Gaussian model, are focused on and presented in more detail than some of the more arcane processes which can be used for modelling. This is because these models are either more widely used or studied. It is important to have a good understanding of Lévy processes on this level, because this is crucial for the next step – simulation. In certain cases the theory is simple and tractable, so that simulation approaches and algorithms follow easily. For other processes the theory is less pleasant and alternate approaches for the purposes of simulation are required. In this case, simulation by means of approximation becomes a very attractive option, since the Lévy measures are generally known in closed form solution. This is an approach which is discussed in Section 5.6.

3.1 The gamma processes

The exponential distribution is frequently used to model the arrival times of objects which arrive at a constant average rate. This is an example of a process which is memoryless. The reason why the exponential distribution is used is as follows: let N_t be the number of objects which have arrived by time t. N_t is clearly a counting process and is non-decreasing in t. Further, assume that the process has independent and stationary increments. Let T be a random variable which represents the time until the first arrival. The set $\{N_t = 0\}$ occurs if and only if $\{t < T\}$ and therefore the two sets have equal measure $\mathbb{P}[N_t = 0] = \mathbb{P}[t < T]$. It can be deduced from the independent and stationary increments of N_t that

$$\begin{aligned}\mathbb{P}[N_{t+s} = 0] &= \mathbb{P}[N_{t+s} - N_t = 0 \cap N_t = 0] \\ &= \mathbb{P}[N_{t+s} - N_t = 0]\mathbb{P}[N_t = 0] \\ &= \mathbb{P}[N_s = 0]\mathbb{P}[N_t = 0]\end{aligned}$$

In other words the functional dependence is multiplicative: $p(t + s) = p(s)p(t)$, which implies that $\mathbb{P}[N_t = 0]$ is modelled by an exponential function (i.e. there exists a λ such that $\mathbb{P}[N_t = 0] = e^{-\lambda t}$. Recall that the exponential distribution is infinitely divisible, and the Lévy process associated with the exponential distribution is the *gamma process*:

Definition 3.1 (Gamma Process). : *Let* $(G_t)_{t\geq 0} = G_t(\alpha, \beta)$ *be a stationary process, with independent increments which are distributed according to the gamma distribution:*

$$f_{G_t(\alpha,\beta)}(x) = \frac{\beta^{\alpha t} x^{\alpha t - 1} e^{-\beta x}}{\Gamma(\alpha t)} \mathbb{I}_{\{x \geq 0\}} \tag{3.1}$$

The stationarity of increments implies that $G_{t+h} - G_t \stackrel{d}{=} G_h$, *for all* $h > 0$. *The gamma process can easily be seen to be a non-decreasing process, and satisfies the conditions for a subordinator.*

Here, the gamma process has been defined as an explicit function of time, manifest through the shape parameter: i.e $G_t(\alpha, \beta) = G(\alpha t, \beta)$. The increments follow a gamma distribution with the following property holding true for all $h > 0$:

$$G_{t+h}(\alpha, \beta) - G_t(\alpha, \beta) \stackrel{d}{=} \mathcal{G}(\alpha^2 h/\beta, \alpha h/\beta) \tag{3.2}$$

where $\mathcal{G}(\alpha, \beta)$ denotes the gamma distribution with parameters α and β. The gamma distribution probability density function is given by:

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x} \tag{3.3}$$

In this form the gamma distribution has the following mean and variance:

$$\mathbb{E}[\mathcal{G}] = \frac{\alpha}{\beta} \tag{3.4}$$

$$\begin{aligned} \mathrm{Var}[\mathcal{G}] &= \mathbb{E}[\mathcal{G}^2] - (\mathbb{E}[\mathcal{G}])^2 \\ &= \frac{\alpha}{\beta^2} \end{aligned} \tag{3.5}$$

It can be verified by taking expectations of (3.1) that:

$$\mathbb{E}[G_t] = \frac{\alpha t}{\beta} \qquad \mathrm{Var}[G_t] = \frac{\alpha t}{\beta^2} \tag{3.6}$$

In using the gamma process to define a stochastic process indexed by time, one would like on average, to 'synchronise' the stochastic clock with the 'real-world' or calendar clock. Thus, a desired property of these gamma processes would be to choose α or β such that $\mathbb{E}[\mathcal{G}_t] = t$. This is possible through simple choice of $\alpha = \beta$. For convenience, define the quantity $\kappa = \alpha^{-1}$. Then (in terms of this parameterisation) one has:

$$\mathbb{E}[G_t] = \frac{\alpha t}{\alpha} = t \qquad \mathrm{Var}[G_t] = \frac{t}{\alpha} = \kappa t \tag{3.7}$$

Sampling happens from a gamma distribution with both shape and scale parameters equal to $1/\kappa$. The process now has mean equal to t, and variance given by the parameter κ. Since the gamma distribution is infinitely divisible, the existence of a Levy process is guaranteed (Proposition 2.1). The distribution of this Lévy process at time 1, X_1, has a gamma distribution.

The gamma process satifies the conditions for a subordinator (Definition 2.10) and can be used to subordinate a Brownian motion. Since subordinators have no negative jumps, the *Laplace exponent* may be used instead of the characteristic exponent. The presence of the Lévy measure in the Laplace exponent means it can be calculated by simply knowing the *moment generating*

function of a gamma distribution. This is given by $(1-\frac{u}{\beta})^{-\alpha}$. By using (2.32) it follows that $\ell(u) = -\alpha \ln(1-u/\beta)$ and then the following can be done [74]:

$$\begin{aligned}
-a\ln(1-\frac{u}{\beta}) &= -\alpha \int_0^u \frac{1}{\beta - y} dy \qquad \text{(for } u \leq 0\text{)} \\
&= -\alpha \int_0^u \int_0^\infty e^{-\beta x + yx} dx dy \\
&= -\alpha \int_0^\infty e^{-\beta x} \int_0^u e^{yx} dx dy \\
&= \alpha \int_0^\infty e^{-\beta x} \frac{1}{x}(e^{ux} - 1) dx \qquad \text{(since } u \leq 0\text{)} \\
&= \int_0^\infty (e^{ux} - 1) \frac{\alpha e^{-\beta x}}{x} dx
\end{aligned} \tag{3.8}$$

This result then shows that the density of the Lévy measure is:

$$\nu(x) = \frac{\alpha e^{-\beta x}}{x} \mathbb{I}_{x>0} = \frac{e^{-x/\kappa}}{\kappa x} \mathbb{I}_{x>0} \tag{3.9}$$

The last equality holds under the parameterisation which calibrates the 'stochastic' clock to coincide with calender time on average. The condition on the Levy measure of a subordinator (that $\int_0^\infty (x \wedge 1)\nu(x)dx < \infty$) is satisfied, but $\int_0^\infty \nu(x)dx = \infty$ meaning that jumps in the gamma process arrive infinitely often. This is because the Lévy measure contains the expected number of jumps, for all jump sizes, arriving per unit time. The gamma process therefore does not necessarily require a martingale component; it is a pure jump process (meaning its movements are entirely through 'jumps') although the majority of these jumps are small (less than 1). The characteristic triplet is completed by calculating γ, given by $\int_0^1 x\nu(x)dx = \int_0^1 \alpha e^{-\beta x} = \alpha(1-e^{-\beta x})/\beta$. Thus, the characteristic triplet of a gamma process is $(A, \gamma, \nu) = (0, \frac{\alpha e^{-\beta x}}{x}, \alpha(1-e^{-\beta x})/\beta)$.

3.2 Inverse Gaussian process

Another subordinating Lévy process is the inverse Gaussian process, which is based on the *inverse Gaussian distribution.* The *inverse Gaussian distribution* describes the distribution of time taken by a Brownian Motion with positive drift to reach a fixed positive level. This distribution is characterised by having two parameters $(\delta, \gamma > 0)$, and density function:

$$f_{IG}(x; \delta, \gamma) = \frac{\delta e^{\delta\gamma}}{\sqrt{2\pi}} x^{-3/2} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right) \mathbb{I}_{x>0} \tag{3.10}$$

An alternate parameterisation in terms of μ and λ is sometimes used. Their equivalence is given by the relations $\delta^2 = \lambda$ and $\gamma^2 = \lambda/\mu^2$. The density then looks as follows:

$$f_{IG}(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) \mathbb{I}_{x>0} \tag{3.11}$$

The mean and variance of the inverse Gaussian distribution are as follows:

$$\mathbb{E}[X_{IG}] = \frac{\delta}{\gamma} = \mu \qquad\qquad \text{Var}[X_{IG}] = \frac{\delta}{\gamma^3} = \frac{\mu^3}{\lambda} \tag{3.12}$$

The inverse Gaussian distribution belongs to the class of *infinitely divisible* distributions. Proposition (2.1) implies that there is a Lévy process associated with this distribution – it is known

as the *normal inverse Gaussian* process. In addition to all of this, the inverse Gaussian distribution is also closed under convolution, meaning that if Z_1 and Z_2 are two independent inverse Gaussian variables with $Z_1 \sim \text{IG}(\delta_1, \gamma)$ and $Z_2 \sim \text{IG}(\delta_2, \gamma)$, then $Z_1 + Z_2 \sim \text{IG}(\delta_1 + \delta_2, \gamma)$. Using the (δ, γ) parameterisation (3.10), the interpretation of the above density is of the first passage time of a Brownian motion, with drift γ, to a level δ [40].

With all this in mind, the *inverse Gaussian process* is now defined:

Definition 3.2 (IG Process). : *Let $(\mathcal{I}_t)_{t\geq 0}(\alpha, \beta)$ be a stationary process, with independent increments which are distributed according to the inverse Gaussian distribution:*

$$f_{\mathcal{I}_t}(x; \delta, \gamma) = \frac{\delta t e^{\delta t \gamma}}{\sqrt{2\pi}} x^{-3/2} \exp\left(-\frac{1}{2}(\delta^2 t^2 x^{-1} + \gamma^2 x)\right) \mathbb{I}_{x>0} \tag{3.13}$$

The stationarity of increments implies that $\mathcal{I}_{t+h} - \mathcal{I}_t \overset{d}{=} \mathcal{I}_h$, for all $h > 0$. As with the gamma process, the inverse Gaussian process satisfies the conditions for a subordinator.

The characteristic function is given by:

$$\phi_{\mathcal{I}_t}(u; \delta, \gamma) = \exp(-\delta(\sqrt{-2iu + \gamma^2} - \gamma)) \tag{3.14}$$

and Lévy measure by,

$$\nu(x) = \frac{\delta e^{-\gamma x}}{x^{3/2}} \mathbb{I}_{x>0}$$

For the inverse Gaussian *distribution* the following scaling property holds true: If $X \sim \mathcal{I}(a, b)$, then $c^2 X \sim \mathcal{I}(ca, b/c)$.

3.3 The variance-gamma process

The first model outside of the eponymous Black-Scholes framework that is treated in detail is the variance-gamma process. This process was introduced in 1990 by Madan and Seneta in symmetric form when they considered a time change of a Brownian motion by means of a driftless gamma process. Madan & Milne [59] then investigated equilibrium option pricing for a symmetric variance-gamma process in a representative agent model with a constant relative risk aversion utility function. The resulting risk-neutral process they obtained ended up being quite similar to the general variance-gamma model proposed in Madan, Carr & Chang [19]. This paper is also important with regards to information concerning instrument valuation.

Since its introduction, the variance-gamma process has gained momentum and usage. The interpretation of the negative drift in a Brownian motion is that it gives rise to positive risk aversion. The motivation for replacing a continuous business time with a 'stochastic clock' whose increments are gamma distributed stems from the fact that exponential distribution is memoryless – what happens in the future does not depend on what has happened up until the present. Mathematically, one writes:

$$\mathbb{P}[X_t > s + t | X_t > s] = \mathbb{P}[X_t > t] \tag{3.15}$$

It can be shown that the only function which possesses this property is the exponential function, hence motivating the use of the *exponential distribution* for modelling inter-arrival times of a whole range of objects. The exponential distribution is a special case of the gamma distribution and in addition to this, the gamma distribution is closed under convolution. Thus, there is a heuristic motivation to use the gamma process for modelling a sequence of jump times which

results in changing the arrival of information to be inherently random.

A pleasant feature of the variance-gamma process, is that the log-normal return density and Black-Scholes formula are special cases[1], making this model an extension of the standard financial modelling paradigm. The variance-gamma model offers certain more 'flexible' advantages over the Black-Scholes model. The variance-gamma model is a 3 parameter stochastic model, having parameters explicitly controlling for kurtosis (symmetric increase in both left and right tail probabilities) and skewness (asymmetry in these left and right tails) properties in the return distribution, in addition to the volatility of the subordinated Brownian motion. The Brownian motion is conditional on a realisation of a random gamma distributed time change – it is subordinated by a gamma process.

3.3.1 Properties of the variance-gamma process

Definition 3.3. (Variance-Gamma process): *Let X_t be a real valued, one dimensional, arithmetic Brownian motion, with drift θ and volatility σ, so that $X(t;\omega) = \theta t + \sigma B(t;\omega)$ and let $G(t;\omega,\alpha,\beta) = G_t(\alpha,\beta)$ be a one dimensional gamma subordinator. Define the process $Y(t;\omega) = X(G_t(\alpha,\beta);\omega) = \theta G_t(\alpha,\beta) + \sigma B(G_t(\alpha,\beta);\omega)$. Y_t is a variance-gamma process.*

The variance of a standard Brownian motion is well known: $\mathrm{Var}(B_t) = t$. The variance-gamma process' name thus originates from the fact that the variance of the Brownian motion in the above definition is given by a gamma process – conditional on the random time change G_t, i.e:

$$\mathrm{Var}(Y_t|G_t) = G_t \tag{3.16}$$

Characteristic Triplet of a variance-gamma process

In Section 3.1, the characteristic triplet of a gamma process was developed. This can now be extended to the variance-gamma process by means of Proposition 2.5. First note that the gamma subordinator used here has no drift and consists purely of jumps. Thus, X_t only jumps when G_t does and therefore Y_t has no points of continuity. By implication, this means that there is no Brownian component and therefore that $\Sigma^Y = 0$, which can also be verified by means of equation (2.36a). The Lévy measure is calculated as follows:

$$\begin{aligned}
\nu^Y &= \int_0^\infty f(x,s)\nu(s)ds \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 s}} \exp\left(-\frac{(x-\theta s)^2}{2\sigma^2 s}\right) \frac{e^{-s/\kappa}}{\kappa s} ds \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 s}} \exp\left[-\frac{x^2}{2\sigma^2 s} - \frac{\theta^2 s}{2\sigma^2} + \frac{x\theta}{\sigma^2}\right] \frac{e^{-s/\kappa}}{\kappa s} ds \\
&= \frac{\exp(x\theta/\sigma^2)}{\kappa} \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 s}} \exp\left[-\frac{x^2}{2\sigma^2} s^{-1} - \frac{\theta^2}{2\sigma^2} s\right] \frac{e^{-s/\kappa}}{s} ds \\
&= \frac{e^{\frac{x\theta}{\sigma^2}} |x| \exp\left[-\frac{|x|\sqrt{\theta^2+2\sigma^2/\kappa}}{\sigma^2}\right]}{\kappa|x| \exp\left[-\frac{|x|\sqrt{\theta^2+2\sigma^2/\kappa}}{\sigma^2}\right]} \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 s^3}} \exp\left[-\frac{x^2}{2\sigma^2} s^{-1} - \left(\frac{\theta^2}{2\sigma^2} + \frac{1}{\kappa}\right) s\right] ds \\
&= \ldots \int_0^\infty \frac{|x| s^{-3/2}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{x^2}{2\sigma^2} s^{-1} - \left(\frac{\theta^2}{2\sigma^2} + \frac{1}{\kappa}\right) s + \frac{|x|\sqrt{\theta^2 + \frac{2\sigma^2}{\kappa}}}{\sigma^2}\right] ds
\end{aligned} \tag{3.17}$$

[1] shown in section on characteristic function and moments of the variance-gamma process

The term in front of the integral is a constant is ignored for now. Evaluation of the integral is more interesting and is, fortunately, simple. It evaluates to 1 because it corresponds to the density of an inverse Gaussian distribution (3.13). The integral in (3.17) follows by means of the following substitutions to the parameters in (3.13) :

$$\delta t = \frac{|x|}{\sqrt{\sigma^2}}, \qquad \gamma^2 = 2\left(\frac{\theta^2}{2\sigma^2} + \frac{1}{\kappa}\right) \tag{3.18}$$

Thus, the Lévy measure is:

$$\nu^Y(x) = \frac{\exp\left[\frac{x\theta}{\sigma^2}\right]\exp\left[-\frac{|x|\sqrt{\theta^2+2\sigma^2/\kappa}}{\sigma^2}\right]}{\kappa|x|} \tag{3.19}$$

To calculate the 'drift' term, one uses:

$$\begin{aligned}
\gamma^Y &= \int_0^\infty \frac{e^{-\kappa s}}{\kappa s}\int_{|x|\leq 1}\frac{x}{\sqrt{2\pi\sigma^2 s}}e^{-\frac{(x-\theta s)^2}{2\sigma^2}}dxds \\
&= \int_{|x|\leq 1} x\int_0^\infty \frac{e^{-\kappa s}}{\kappa s}\frac{1}{\sqrt{2\pi\sigma^2 s}}e^{-\frac{(x-\theta s)^2}{2\sigma^2}}dsdx \\
&= \int_{|x|\leq 1} x\nu^Y dx
\end{aligned} \tag{3.20}$$

Thus, the characteristic triplet of a variance-gamma process is given by:

$$\left(0, \frac{\exp\{Ax - B|x|\}}{\kappa|x|}, \int_{|x|\leq 1} x\frac{\exp\{Ax - B|x|\}}{\kappa|x|}dx\right) \tag{3.21}$$

where,

$$A = \frac{\theta}{\sigma^2}, \qquad B = \frac{\sqrt{\theta^2 + 2\sigma^2/\kappa}}{\sigma^2}$$

Characteristic Function and Moments of a variance-gamma process

The characteristic function of a variance-gamma process, $VG(\theta, \sigma, \kappa)$, can be obtained by some rather cumbersome integration and the use of the Lévy -Khinchin representation theorem. An easier method, involves using conditional expectations as follows (remember that $\kappa = \alpha^{-1}$):

$$\begin{aligned}
\mathbb{E}[e^{iuY_t}] &= \mathbb{E}[\mathbb{E}[e^{iuY_t}|G(t;\omega) = z]] \\
&= \mathbb{E}[\mathbb{E}[e^{iu(\theta z + \sigma B_z)}|G(t;\omega) = z]] \\
&= \mathbb{E}[e^{iu\theta z}\mathbb{E}[e^{iu\sigma B_z}|G(t;\omega) = z]] \\
&= \mathbb{E}[e^{iu\theta z}e^{-u^2\sigma^2 z/2}|G(t;\omega) = z] \\
&= \mathbb{E}[e^{z(iu\theta - u^2\sigma^2/2)}|G(t;\omega) = z] \\
&= \mathbb{E}[e^{G_t(iu\theta - u^2\sigma^2/2)}] \\
&= \left(1 - \kappa(i\theta u - \tfrac{1}{2}\sigma^2 u^2)\right)^{-t/\kappa}
\end{aligned} \tag{3.22}$$

where the last line follows from the definition of the moment generating function of a gamma distribution with an argument of $(iu\theta - u^2\sigma^2/2)$. The *characteristic exponent* is then:

$$\Psi(u) = -\frac{1}{\kappa}\log(1 + \frac{u^2\sigma^2\kappa}{2} - i\theta\kappa u) \tag{3.23}$$

To see that the lognormal density is a special case of the variance-gamma process, first observe that under symmetric ($\theta = 0$) variance-gamma dynamics (3.22) reduces to

$$\mathbb{E}[e^{iuY_t}] = (1 + \tfrac{1}{2}\sigma^2 u^2 \kappa)^{-t/\kappa} \tag{3.24}$$

Letting $\kappa \to 0^+$, the above expression can easily be seen to converge to an exponential function:

$$\mathbb{E}[e^{iuY_t}] = \exp[-\tfrac{1}{2}\sigma^2 u^2] \tag{3.25}$$

which happens to be the well known characteristic function of a normal distribution.

□

The first four central moments of the variance-gamma process are given as follows [19]:

$$\mu_1 = \theta t \tag{3.26a}$$
$$\mu_2 = t(\theta^2\kappa + \sigma^2) \tag{3.26b}$$
$$\mu_3 = t\kappa\theta(3\sigma^2 + 2\theta^2\kappa) \tag{3.26c}$$
$$\begin{aligned}\mu_4 &= t(3\sigma^4\kappa + 6\theta^4\kappa^3 + 12\sigma^2\theta^2\kappa^2) + t^2(3\sigma^4 + 6\sigma^2\theta^2\kappa + 3\theta^4\kappa^2) \\ &= t\kappa(3\sigma^4 + 6\theta^4\kappa^2 + 12\sigma^2\theta^2\kappa) + 3t^2(\theta^2\kappa + \sigma^2)^2\end{aligned} \tag{3.26d}$$

These four moments relate to four key distributional properties in statistics. The first non-central moment (μ_1) gives the mean, or average value at a specific time point (t) and the variance is given by (μ_2). The third and fourth moments are the *skewness*($= \varsigma$) and *kurtosis*[2] ($= \kappa$), which measure the symmetry and difference in curvature to the normal distribution. It is often easier to work with these quantities in coefficient (or normalised) form, which are defined as:

$$\varsigma(X_t) = \frac{\mu_3}{\sigma^3} \qquad\qquad \kappa(X_t) = \frac{\mu_4}{\sigma^4} \tag{3.27}$$

The Gaussian distribution has a kurtosis coefficient of 3 and a skewness coefficient of 0. This provides a benchmark for comparison with other distributions. For path simulation, theoretical moments also provide a means of comparison for purposes of accuracy. It is a straightforward matter to calculate the sample moments of a series of values for a given time point. A check can then be carried out to assess whether or not the calculated moments are in good agreement with the theoretical moments.

The variance-gamma process can be described as a normal variance-mean mixture, where the mixing density is gamma. This is equivalent to saying that conditional on any gamma time change $G_t = g$, the variance-gamma variate, $X(t)$, over an interval of length t is normally distributed, with mean θg and standard deviation $\sigma\sqrt{g}$. X may thus be written as:

$$X(t)|_{G_t=g} = \theta g + \sigma\sqrt{g}Z \tag{3.28}$$

where $Z \sim N(0,1)$ independent of g. Proof that relationships (3.26) hold can be seen using the above conditional distribution assumptions of a variance-gamma process. For example, for the first moment:

$$\begin{aligned}\mathbb{E}[Y_t|G_t = g] &= \mathbb{E}[\theta g + \sigma\sqrt{g}X] \\ &= \theta\mathbb{E}[g] + \sigma\mathbb{E}[\sqrt{g}]\mathbb{E}[X] \\ &= \theta t + 0 = \theta t\end{aligned} \tag{3.29}$$

[2] some textbooks and programs often work with the quantity called the *excess kurtosis*, which is defined as $\kappa - 3$. It is important to know which quantity is being used by the textbook or computer program

since $\mathbb{E}[G_t] = t$ and $\mathbb{E}[X] = 0$. One can now define the central quantity

$$L_t = G_t - \mathbb{E}[G_t] = \theta(g - t) + \sigma\sqrt{g}X \tag{3.30}$$

Then, for example, the variance is calculated as $\mathbb{E}[L_t^2]$. This yields:

$$\begin{aligned} Var[G_t] &= \mathbb{E}[(\theta(g-t) + \sigma\sqrt{g}X)^2] \\ &= \mathbb{E}[\theta^2(g-t)^2 + 2\theta(g-t)\sigma\sqrt{g}X + \sigma^2 g X^2] \\ &= \theta^2\mathbb{E}[(g-t)^2] + 2\theta\sigma\mathbb{E}[\sqrt{g}(g-t)]\mathbb{E}[X] + \sigma^2\mathbb{E}[g]\mathbb{E}[X^2] \\ &= \theta^2\kappa t + 0 + \sigma^2 t \cdot 1 \\ &= t(\sigma^2 + \theta^2\kappa) \end{aligned} \tag{3.31}$$

The additional proofs can be found in [19].

Probability density function of a variance-gamma process

The probability density function of a variance-gamma process is Gaussian conditional on realisation of $G(t;\omega) = g$ following a gamma distribution with parameters t/κ and $1/\kappa$. Thus:

$$f_{X(t)|G(t)=g}(x) = \frac{1}{\sqrt{2\pi\sigma^2 g}} \exp\left(-\frac{(x-\theta g)^2}{2\sigma^2 g}\right) \tag{3.32}$$

$$\begin{aligned} f_{X(t)} &= \int_0^\infty f_{X(t)|G(t)=g}(x) f_{G(t)} dg \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 g}} \exp\left(-\frac{(x-\theta g)^2}{2\sigma^2 g}\right) \frac{g^{\frac{t}{\kappa}-1} e^{-g/\kappa}}{\kappa^{\frac{t}{\kappa}}\Gamma\left(\frac{t}{\kappa}\right)} dg \end{aligned} \tag{3.33}$$

The middle line simply states that the unconditional density of $f_{X(t)}$ is the integral of the product of the conditional density and the marginal density over the domain of the marginal density. This can be fleshed out and simplified in terms of special mathematical functions, as shown below, to obtain a 'simplified' form:

$$= \frac{1}{\sqrt{2\pi\sigma^2}\Gamma\left(\frac{t}{\kappa}\right)\kappa^{\frac{t}{\kappa}}} \int_0^\infty g^{\frac{t}{\kappa}-\frac{1}{2}-1} \exp\left[-\frac{x^2}{2\sigma^2 g} + \frac{x\theta}{\sigma^2} - \frac{g\theta^2}{2\sigma^2} - \frac{g}{\kappa}\right] dg \tag{3.34}$$

$$= \frac{\exp\left\{\frac{x\theta}{\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}\Gamma\left(\frac{t}{\kappa}\right)\kappa^{\frac{t}{\kappa}}} \int_0^\infty g^{(\frac{t}{\kappa}-\frac{1}{2})-1} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2 g} + g\left\{\frac{\theta^2}{\sigma^2} + \frac{2}{\kappa}\right\}\right)\right] dg \tag{3.35}$$

At this point, progress is contingent on a non-obvious substitution: Let

$$p = \frac{\sqrt{2\sigma^2/\kappa + \theta^2}}{x} g \tag{3.36}$$

This, amongst other things, means that the following relations hold:

$$dp = \frac{\sqrt{2\sigma^2/\kappa + \theta^2}}{x} dg \qquad \frac{x^2}{\sigma^2 g} = \frac{\sqrt{2\sigma^2/\kappa + \theta^2}x}{p\sigma^2}$$
$$g = \frac{xp}{\sqrt{2\sigma^2/\kappa + \theta^2}} \tag{3.37}$$

The focus at this point is solely on the exponential term and the way in which it simplifies. At present one has:

$$\begin{aligned}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2 g}+g\left\{\frac{\theta^2}{\sigma^2}+\frac{2}{\kappa})\right\}\right)\right] &= \exp\left[-\frac{1}{2}\left(\frac{\sqrt{\frac{2\sigma^2}{\kappa}+\theta^2}x}{p\sigma^2}+\frac{xp}{\sqrt{2\sigma^2/\kappa+\theta^2}}\left\{\frac{\theta^2}{\sigma^2}+\frac{2}{\kappa}\right\}\right)\right] \\ &= \exp\left[-\frac{x}{2}\sqrt{\frac{2\sigma^2}{\kappa}+\theta^2}\left(\frac{1}{p\sigma^2}+\frac{p}{(\frac{2\sigma^2+\theta^2\kappa}{\kappa})}\frac{(2\sigma^2+\theta^2\kappa)}{\sigma^2\kappa}\right)\right] \\ &= \exp\left[-\frac{x\sqrt{\frac{2\sigma^2}{\kappa}+\theta^2}}{2\sigma^2}\left(\frac{1}{p}+p\right)\right] \qquad (3.38)\end{aligned}$$

The $g^{\frac{t}{\kappa}-3/2}dg$ term is manipulated using the differential relations obtained in (3.37):

$$\begin{aligned}g^{(\frac{t}{\kappa}-\frac{1}{2})-1}dg &= \left(\frac{xp}{\sqrt{2\sigma^2/\kappa+\theta^2}}\right)^{\frac{t}{\kappa}-\frac{3}{2}}\frac{xdp}{\sqrt{2\sigma^2/\kappa+\theta^2}} \\ &= \left(\frac{1}{2\sigma^2/\kappa+\theta^2}\right)^{\frac{t}{2\kappa}-\frac{3}{4}+\frac{1}{2}} x^{\frac{t}{\kappa}-\frac{1}{2}}p^{\frac{t}{\kappa}-\frac{3}{2}}\,dp \qquad (3.39) \\ &= \left(\frac{x^2}{2\sigma^2/\kappa+\theta^2}\right)^{\frac{t}{2\kappa}-\frac{1}{4}} p^{\frac{t}{\kappa}-\frac{3}{2}}\,dp \qquad (3.40)\end{aligned}$$

Everything must be brought together to obtain an integral representation for the probability density function:

$$f_X(x)=\frac{2\exp\left[\frac{x\theta}{\sigma^2}\right]}{\kappa^{\frac{t}{\kappa}}\Gamma(\frac{t}{\kappa})\sigma\sqrt{2\pi}}\left(\frac{x^2}{2\sigma^2/\kappa+\theta^2}\right)^{\frac{t}{2\kappa}-\frac{1}{4}}\frac{1}{2}\int_0^\infty \exp\left[-\frac{1}{2}\frac{x\sqrt{\frac{2\sigma^2}{\kappa}+\theta^2}}{\sigma^2}\left(\frac{1}{p}+p\right)\right]p^{(\frac{t}{\kappa}-\frac{1}{2})-1}\,dp \qquad (3.41)$$

Introducing the modified Bessel function of the 3^{rd} kind, K_φ, allows for further simplification. These functions have integral representations:

$$K_\varphi(\varsigma)=\frac{1}{2}\int_0^\infty t^{\varphi-1}\exp\left[-\frac{\varsigma}{2}\left(\frac{1}{t}+t\right)\right]dt \qquad (3.42)$$

On recognising this to be the integral in the above expression with parameter values given by:

$$\varsigma=\frac{\sqrt{x^2(2\sigma^2/\kappa+\theta^2)}}{\sigma^2}, \qquad \varphi=\frac{t}{\kappa}-\frac{1}{2}, \qquad p=t \qquad (3.43)$$

the variance-gamma pdf can be represented as follows:

$$\begin{aligned}f_X(x) &= \sqrt{\frac{2}{\pi}}\frac{\exp\left[\frac{x\theta}{\sigma^2}\right]}{\kappa^{\frac{t}{\kappa}}\Gamma(\frac{t}{\kappa})\sigma}\left(\frac{x^2}{2\sigma^2/\kappa+\theta^2}\right)^{\frac{\varphi}{2}}K_\varphi(\varsigma) \qquad (3.44) \\ &= C\exp[Ax]\left(\frac{|x|}{B\sigma^2}\right)^{\varphi}K_\varphi(\varsigma) \qquad (3.45)\end{aligned}$$

Note that for the value $x=0$ the argument of the Bessel function is also 0, and since around the origin, these Bessel functions behave like x^{-n} [2], for this particular value the pdf explodes.

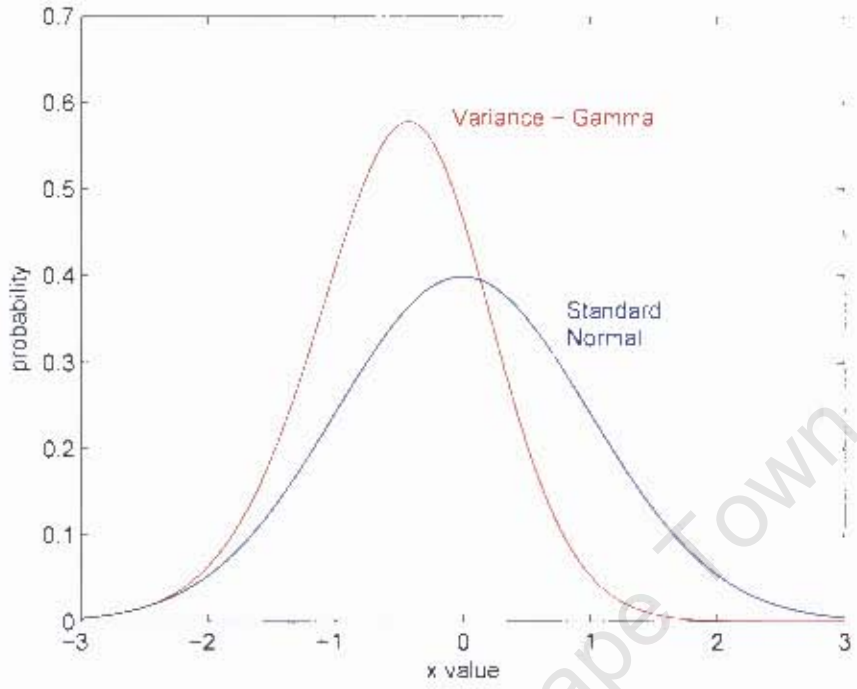

FIGURE 3.1: Plot illustrating major differences between the variance-gamma process and the Gaussian distribution

The discontinuity, however, is removable (as can be seen graphically) and therefore does not pose any numerical and computational problems as far as integration goes. Figure 3.1 shows the probability density function of a variance-gamma process (red) with parameters $\theta = \quad 0.5, \sigma = 0.7$ and $\kappa = 0.1$. A standard Gaussian distribution has been overlayed, drawing attention to the major differences between using the standard normal distribution and a fat-tailed, asymmetric and leptokurtic distribution to model asset returns.

An alternate representation

The sum of the absolute log changes is finite for the variance-gamma process although it is a well known fact that the same is not true for Brownian motion. This implies that the variance-gamma process is of finite variation (using a standard result from analysis) which means that it can be written as the difference of two increasing (gamma) processes. This can be seen by looking at the characteristic function (3.22). It can be factorised into two brackets as follows:

$$\mathbb{E}[e^{iuY_t}] = \frac{1}{(1 - iu\eta_p)^{t/\kappa}} \cdot \frac{1}{(1 + iu\eta_n)^{t/\kappa}} \tag{3.46a}$$

$$= \frac{1}{(1 - iu\eta_p)^{t/\kappa}} \cdot \frac{1}{(1 - iu\eta_n)^{t/\kappa}} \tag{3.46b}$$

where η_p, η_n have yet to be quantified. In general the characteristic function of a *gamma process* is given as:

$$\mathbb{E}[e^{iuG_t}] - \phi_G = \frac{1}{(1 - iu/\beta)^{\alpha t}} \tag{3.47}$$

This particular gamma distribution has a *mean parameter* $= \alpha/\beta$ and variance parameter $= \alpha/\beta^2$ as noted in Section 3.1. For simplicity, call the mean parameter μ and the variance

parameter ν. The characteristic function should then look as follows:

$$\mathbb{E}[e^{iuG_t}] = \left(\frac{1}{1 - iu\left(\frac{\nu}{\mu}\right)}\right)^{\left(\frac{\mu^2}{\nu}\right)t} \tag{3.48}$$

The μ and ν parameters still need to be defined from the characteristic function of the variance-gamma process. This is done as follows: The first term in (3.46b) is the characteristic function of a gamma process with $\mu_p/\nu_p = \eta_p$ and $\mu_p^2/\nu_p = 1/\kappa$. The second term is the complex conjugate of the characteristic function of a gamma process with $\mu_n/\nu_n = \eta_n$ and $\mu_n^2/\nu_n = 1/\kappa$. Denoting the first gamma function as γ_p and the second as γ_n, (3.46b) can be expressed as

$$\begin{aligned}
\mathbb{E}[e^{iuY_t}] &= \phi_{\gamma_n}(u) \cdot \overline{\phi_{\gamma_p}(u)} && (3.49a)\\
&= \phi_{\gamma_n}(u) \cdot \phi_{-\gamma_p}(u) && (3.49b)\\
&= \mathbb{E}[e^{iu(\gamma_{nt} - \gamma_{pt})}] && (3.49c)
\end{aligned}$$

illustrating that $\gamma_{pt} - \gamma_{nt} \overset{d}{=} Y_t$. To ensure that equality holds between (3.46a) and (3.22), it is a requirement that:

$$\eta_p\eta_n = \frac{\nu_p}{\mu_p}\frac{\nu_n}{\mu_n} = \frac{1}{2}\sigma^2\kappa \qquad\qquad \eta_p - \eta_n = \frac{\nu_p}{\mu_p} - \frac{\nu_n}{\mu_n} = \theta\kappa \tag{3.50}$$

and, subject to

$$\frac{\mu_n^2}{\nu_n} = \frac{\mu_p^2}{\nu_p} = \frac{1}{\kappa} \tag{3.51a}$$

the following solutions can be found for the quantities $\mu_p, \mu_n, \nu_p, \nu_n$ in terms of θ, σ and κ [19]:

$$\frac{\nu_p}{\mu_p} = \frac{1}{2}\sqrt{\theta^2\kappa^2 + 2\sigma^2\kappa} + \frac{\theta\kappa}{2} \tag{3.51b}$$

$$\frac{\nu_n}{\mu_n} = \frac{1}{2}\sqrt{\theta^2\kappa^2 + 2\sigma^2\kappa} - \frac{\theta\kappa}{2} \tag{3.51c}$$

If the parameters θ, σ and κ are specified as inputs, the variance-gamma process corresponding to a Brownian subordination with drift θ and volatility σ can be represented by the above two gamma processes, and thus by simulating these two gamma processes, sample paths of the variance-gamma process can be constructed.

Using the fact that the variance-gamma process is equivalent in distribution to the difference of the above two gamma processes, an alternate parameterisation of the Lévy measure can be obtained:

$$\nu(dx) = \begin{cases} \frac{C\exp(Gx)}{|x|}dx & (x < 0) \\ \frac{C\exp(-Mx)}{x}dx & (x > 0) \end{cases} \tag{3.52}$$

where,

$$C = \kappa^{-1} \tag{3.53}$$

$$G = \left(\sqrt{\frac{1}{4}\theta^2\kappa^2 + \frac{1}{2}\sigma^2\kappa} - \frac{\theta\kappa}{2}\right)^{-1} \tag{3.54}$$

$$M = \left(\sqrt{\frac{1}{4}\theta^2\kappa^2 + \frac{1}{2}\sigma^2\kappa} + \frac{\theta\kappa}{2}\right)^{-1} \tag{3.55}$$

These parameters have the same starting letters as the surnames of Carr, Geman and Madan, who (along with Marc Yor) went on to introduce a generalisation to this process, the CGMY process.

3.4 The normal inverse Gaussian process

The normal inverse Gaussian process is a process built by subordinating a Brownian motion with an inverse Gaussian subordinator. The process was introduced by Barndorff-Nielsen [10] in 1995 and has since then been studied by himself (see [9, 8] for example) and others (including Rydberg [72], Eberlein et. al [30, 31, 32]). It is a process which is known as a *normal variance-mean mixture where the mixing density is the inverse Gaussian distribution.* This means that the process has normally distributed increments, conditional on an inverse Gaussian time change. The normal inverse Gaussian distribution typically has 4 parameters $(\alpha, \beta, \mu, \delta)$ [9] and the density function is given by (3.59). The normal inverse Gaussian process can also be viewed as a superposition of weighted independent Poisson processes, with weights of all sizes occurring, but mostly dominated by weights which are 'numerically small' [9]. The normal inverse Gaussian (NIG) process can also be defined in terms of parameters θ, σ and κ as for the variance-gamma process. This is because the inverse Gaussian process is also a subordinator. The NIG process is defined from this point of view:

Definition 3.4 (Normal inverse Gaussian process)**.** *Let X_t be a real valued, one dimensional, arithmetic Brownian motion, with drift θ and volatility σ, so that $X(t;\omega) = \theta t + \sigma B(t;\omega)$ and let $\mathcal{I}(t;\omega,\alpha,\beta) = \mathcal{I}_t(\alpha,\beta)$ be a one dimensional inverse Gaussian subordinator. Define the process $Y(t;\omega) = X(\mathcal{I}_t(\alpha,\beta);\omega) = \theta\mathcal{I}_t(\alpha,\beta) + \sigma B(\mathcal{I}_t(\alpha,\beta);\omega)$. Y_t is a normal inverse Gaussian process.*

A derivation of the probability density function, as with the variance-gamma process, follows from using the conditional Gaussian representation of the NIG process. Thus, the process is described by the same parameters as the variance-gamma process. Since the conditional process, $X_{t|\mathcal{I}_t=y} = \theta y + \sigma\sqrt{y}Z$, the *conditional density function* of this 'calibrated' process $(X_t)|\mathcal{I}(t) = y$ would be given by:

$$f_{X(t)|\mathcal{I}(t)=y}(x) = \frac{1}{\sqrt{2\pi y}\sigma}\exp\left[-\frac{(x-\theta y)^2}{2\sigma^2 y}\right] \tag{3.56}$$

The unconditional density (in terms of x) is then obtained by integrating out the y variable:

$$\begin{aligned} f_{X(t)}(x) &= \int_0^\infty f_{X(t)|\mathcal{I}(t)=y} f_{Y(t)}(y)dy \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi y}\sigma}\exp\left[-\frac{(x-\theta y)^2}{2\sigma^2 y}\right]\cdot\frac{te^{t/\kappa}}{\sqrt{2\pi\kappa}}y^{-3/2}\exp\left[-\frac{1}{2\kappa}(\frac{t^2}{y}+y)\right]dy \end{aligned} \tag{3.57}$$

The calculation methodology is similar to the variance-gamma case. Notice that the exponential term under the integral contains a term of the form $(y^{-1}+y)$. By shuffling terms around, taking out constant terms and grouping the others correctly, an intelligent substitution similar to (3.36) can be made. This is done because the $(y^{-1}+y)$ term is present in the modified Bessel functions (of the second kind[3]), whose integral representation is given by (3.42). To uncomplicate the integral, the following substitution is required:

$$p(y) = \sqrt{\frac{\theta^2+\frac{\sigma^2}{\kappa}}{x^2+\frac{t^2\sigma^2}{\kappa}}}\,y \tag{3.58}$$

The probability density function of the NIG process (after simplification) looks as follows,

$$f_{X(t)}(x) = \frac{t}{\pi}\cdot\sqrt{\frac{\theta^2}{\kappa\sigma^2}+\frac{1}{\kappa}}\cdot\frac{K_{-1}\left(\sqrt{x^2+t^2\sigma^2/\kappa}\right)}{\sqrt{x^2+t^2\sigma^2/\kappa}}\exp\{t/\kappa+\beta x\} \tag{3.59}$$

[3] used interchangeably in the literature with the phrase: 'modified Bessel function of the third kind'

after the following transformations[4] of the basic parameters:

$$\delta \to \frac{t}{\sqrt{\kappa}} \qquad \gamma \to \frac{1}{\sqrt{\kappa}} \tag{3.60}$$

$$\alpha \to \frac{1}{\sigma^2}\sqrt{\theta^2 + \sigma^2/\kappa} \qquad \beta \to \theta/\sigma^2 \tag{3.61}$$

The use of these transformations can be used to tidy up (3.59) and obtain the expression given by Barndorff-Nielsen [9]:

$$g^{NIG}(x;\alpha,\beta,\delta) = \alpha\frac{\delta K_1\left(\alpha\sqrt{\delta^2+x^2}\right)}{\pi\sqrt{\delta^2+x^2}}\exp\{\beta x + \delta\gamma\} \tag{3.62}$$

The first bit of (3.62) is the normalising constant for the density. One nice feature of (3.62) is that it provides a simple interpretation of the nature of each parameter: α is the *shape parameter*, $\mu(=0$ in this case) is the *location parameter*, β is an *asymmetry parameter* and δ is a *scale parameter*. The parameters satisfy $0 \leq |\beta| \leq \alpha$, $\mu \in \mathbb{R}$ and $\delta > 0$. Authors such as Cont & Tankov [25, 79] typically work in terms of the θ, σ, κ parameterisation, while Barndorff-Nielsen [9] and others typically work in terms of $(\alpha, \beta, \mu, \delta)$ parameters. Conditional normality means that increments of X over an interval of length t, are normally distributed, with mean θy and variance $\sigma\sqrt{y}$. Thus,

$$X(t)_{|\mathcal{I}_t = y} = \theta y + \sigma\sqrt{y}Z \tag{3.63}$$

where $Z \sim N(0,1)$, $\mathbb{E}[y] = t$ and $\mathrm{Var}[y] = t/\delta^2$. Z is independent of y. From this, the cumulants (or more importantly the central moments) of the NIG process can be calculated. Taking (conditional) expectation yields:

$$\mathbb{E}[X_t|\mathcal{I}_t = y] = \theta t \tag{3.64}$$

as with the variance-gamma process, and this enables one to calculate central moments of the NIG process, given in (3.65).

Under the (θ, σ, κ) parameterisation, these cumulants are [25]:

$$c_1 = \mathbb{E}[X_t] = \theta t \qquad c_2 = \mathrm{Var}[X_t] = t(\sigma^2 + \theta^2\kappa) \tag{3.65a}$$

$$c_3 = 3\theta\kappa t(\sigma^2 + \theta^2\kappa) \qquad c_4 = 3\kappa t(\sigma^4 + 5\theta^4\kappa^2) + 18\sigma^2\theta^2\kappa^2 \tag{3.65b}$$

Central moments relate quite nicely to cumulants, c_n, and it is possible to relate c_3 and c_4 to the skewness and kurtosis of an NIG process. This is done as follows:

$$\varsigma[X_t] = \frac{\mu_3}{\mathrm{Var}[X_t]^{3/2}} = \frac{c_3}{\mathrm{Var}[X_t]^{3/2}} \qquad \kappa[X_t] = \frac{\mu_4}{\mathrm{Var}[X_t]^2} = \frac{c_4 + 3c_2^2}{\mathrm{Var}[X_t]^2} \tag{3.66}$$

Moment Generating Function

The class of normal inverse Gaussian distributions that have fixed values of α and δ form an exponential model with canonical parameter β and x as canonical statistic. This makes it possible to write down the *moment generating function*, as:

$$M(u;\alpha,\beta,\delta) = \exp\left[\delta\{\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+u)^2}\}\right] \tag{3.67}$$

$$= \exp\left[\delta\{\gamma - \sqrt{\alpha^2-(\beta+u)^2}\}\right] \tag{3.68}$$

Therefore the mean, variance, skewness and kurtosis can be defined in terms of these parameters, required when a subordination based Brownian motion approach of simulating a NIG process is not considered.

[4] the transformations of δ and γ are chosen so that $\mathbb{E}[\mathcal{I}_t] = t$ and $\mathrm{Var}[\mathcal{I}_t] = \kappa t$

Characteristic Exponent

The *characteristic exponent* is given by [25]:

$$\Psi(u) = \frac{1}{\kappa}\left(1 - \sqrt{1 + u^2\sigma^2\kappa - 2i\theta u\kappa}\right) \tag{3.69}$$

Lévy measure

The normal inverse Gaussian process has Lévy measure [25]:

$$\nu(x) = \frac{\sqrt{\theta^2 + \sigma^2/\kappa}}{2\pi\sigma\sqrt{\kappa}|x|} \exp\left[\frac{\theta}{\sigma^2}x\right] K_1\left(\frac{\sqrt{\theta^2 + \sigma^2/\kappa}}{\sigma^2}|x|\right) \tag{3.70}$$

Infinite divisibility and limiting cases

The following relation holds for the addition of two independent variables following a NIG distribution:

$$NIG(\alpha, \beta, \mu_1, \delta_1) + NIG(\alpha, \beta, \mu_2, \delta_2) = NIG(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2) \tag{3.71}$$

The NIG distribution is infinitely divisible. It is also the only process in the *general hyperbolic* (see next section) family of distributions that is closed under convolution, even though the latter class still form an infinitely divisible family of distributions. Certain well known distributions can be recovered as special (or limiting) cases of the NIG distribution:
If $\beta = 0$, $\alpha \to \infty$ with $\delta/\alpha \to \sigma^2$ then the distribution tends to a normal distribution $N(\mu, \sigma^2)$. The Cauchy distribution is the special case NIG(0,0,1,0).

3.5 The generalised inverse Gaussian process

The inverse Gaussian distribution (3.13) can be generalised to what is known as the *generalised inverse Gaussian* distribution. The additional or 'generalising' parameter is given by $\lambda \in \mathbb{R}$ and the density function looks as follows [76]:

$$f_{GIG}(x; \lambda, \delta, \gamma) = \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\left[-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right] \mathbb{I}_{x>0} \tag{3.72}$$

Both δ and γ must be non-negative and cannot simultaneously both be equal to 0. The characteristic function is:

$$\phi_{GIG}(u; \lambda, \delta, \gamma) = \frac{1}{K_\lambda(\delta\gamma)}(1 - 2iu/b^2)^{\lambda/2} K_\lambda(\delta\gamma\sqrt{1 - 2iub^{-2}}) \tag{3.73}$$

where K_λ is the modified Bessel function of the second kind with index λ. Infinite divisibility of this distribution was shown by Barndorff-Nielsen and Halgreen [11]. Use of process of this type dates as far back as 1953 when I. J. Good used them to study the population frequencies of certain species [41]. A standard reference work for the generalised inverse Gaussian distribution is Jørgenson [53]. The generalised inverse Gaussian distribution is used extensively in geostatistics and the distribution was popularised by Barndorff-Nielsen. It is also known as the Sichel distribution, after Dr. Herbert Sichel one of the pioneers of this distribution. Both the inverse Gaussian and gamma distributions are special cases of the generalised inverse Gaussian distribution [76].

3.6 The generalised hyperbolic family of distributions

The normal inverse Gaussian distribution is nested in a larger class of distributions known as the class of 'general hyperbolic' distributions. These distributions in turn, belong to an even bigger class known as extended generalised Γ-convolutions, which will not be elaborated on any further. Barndorff-Nielsen [8] derived the generalised hyperbolic distribution as a mean-variance mixture of the normal distribution and the generalised inverse Gaussian distribution. The distribution was introduced as a model for the grain size distribution of wind blown sand. Subclasses of this process used for financial modelling are the Hyperbolic distribution (Eblerlein & Keller [29]) and the normal inverse Gaussian process (introduced by Barndorff-Nielsen in 1995). It is used in economics in particular for risk management applications as well as the modelling of financial markets. Its probability density function is given by

$$h_{GH}(x;\lambda,\alpha,\beta,\delta,\mu) = a(\lambda,\alpha,\beta,\delta,\mu)(\delta^2+(x-\mu)^2)^{(\lambda-\frac{1}{2})/2}e^{\beta(x-\mu)} \times K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) \tag{3.74}$$

where the normalisation constant is given as:

$$a(\lambda,\alpha,\beta,\delta,\mu) = \frac{(\alpha^2-\beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^{\lambda}K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})} \tag{3.75}$$

The parameters can be interpreted in the same way as they were for the normal inverse Gaussian distribution. The additional parameter $\lambda \in \mathbb{R}$ plays a role in the classification of certain subclasses while having considerable influence over the concentration of mass in the tails. The characteristic function is [76]

$$\varphi(u) = e^{i\mu u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+iu)^2}\right)^{\lambda/2}\frac{K_{\lambda}(\delta\sqrt{\lambda^2-(\beta iu)^2})}{K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})} \tag{3.76}$$

Barndorff-Nielsen [9], claims that the normal inverse Gaussian distribution can approximate most general hyperbolic processes reasonably well and has the additional feature of being able to describe observations with considerably heavier tail behaviour than that characterising the hyperbolic shape (log linear rate of decrease). Also, the normal inverse Gaussian distribution possesses more tractable probabilistic properties than the hyperbolic distributions and so seems to be a more useful and likely choice for modelling. Special cases and limiting functions include the following distributions [25]:

- When $\delta \to \infty$ and $\delta/\alpha \to \sigma^2$ the normal distribution is recovered.
- For $\lambda = 1$, the process corresponds to the hyperbolic distribution. For this distribution, the logarithm of its density is a hyperbola (the same is true for the GH process) – hence the name.
- The normal inverse Gaussian process is the special case $\lambda = -1/2$, while the variance-gamma process is given when $\delta = 0$ and $\mu = 0$.
- Student t distributions are obtained when $\lambda < 0$, and $\alpha = \beta = \mu = 0$

3.7 α-stable Lévy Processes

Stable processes are those which satisfy the following property:

Definition 3.5 (Stable distribution). A random variable $X \in \mathbb{R}^d$ is said to have a stable distribution if for every $a > 0$ there exists a $b(a) > 0$ and $c(a) \in \mathbb{R}^d$ such that:

$$\Phi_X(z)^a = \Phi_X(zb(a))e^{ic.z} \tag{3.77}$$

It has *strictly stable* distribution if

$$\Phi_X(z)^a = \Phi_x(zb(a)) \tag{3.78}$$

Definition 3.6 (Selfsimilar Lévy processes). A Lévy process is said to be selfsimilar if

$$\forall a > 0, \qquad \left(\frac{X_{at}}{a^{1/\alpha}}\right)_{t\geq 0} \stackrel{d}{=} (X_t)_{t\geq 0} \tag{3.79}$$

The Lévy measure of a stable process is generally given as follows:

$$\nu(x) = \frac{A}{x^{\alpha+1}}\mathbb{I}_{x>0} + \frac{B}{|x|^{\alpha+1}}\mathbb{I}_{x<0} \tag{3.80}$$

The characteristic function has the following form:

$$\Phi_S(z) = \begin{cases} \exp\left\{-\sigma^\alpha|z|^\alpha(1 - i\beta\text{sgn}(z)\tan\frac{\pi\alpha}{2}) + iuz)\right\}, & \alpha \neq 1 \\ \exp\left\{-\sigma|z|(1 + i\beta\frac{2}{\pi}\text{sgn}(z)\log|z|) + iuz)\right\}, & \alpha = 1 \end{cases} \tag{3.81}$$

where $\alpha \in (0, 2]$, $\sigma \geq 0$, $\beta \in [-1, 1]$ and $\mu \in \mathbb{R}$. A stable distribution in this parameterisation is denoted by $S_\alpha(\sigma, \beta, \nu)$. Here, σ is the scale parameter[5], μ is the shift parameter[6] and β is its skewness. When both $\beta = 0$ and $\mu = 0$, X is said to have a symmetric stable distribution, in which case the characteristic function is given by:

$$\Psi_X(z) = \exp(-\sigma^\alpha|z|^\alpha) \tag{3.82}$$

From (3.80) it is possible to realise that α-stable distributions on $\mathbb{R}$ never admit a second moment, and only admit a first moment if $\alpha > 1$. There are only three cases in which the density of an $\alpha-$stable law is known:

- $S_2(\sigma, 0, \mu)$: The Gaussian distribution with law (symmetric around mean and in nonstandard parameterisation)
$$\frac{1}{2\sigma\sqrt{\pi}}e^{-(x-\mu)^2/4\sigma^2} \tag{3.83}$$
- $S_1(\sigma, 0, \mu)$: The Cauchy distribution, symmetric around mean, with law
$$\frac{\sigma}{\pi((x-\mu)^2 + \sigma^2)} \tag{3.84}$$
- $S_{1/2}(\sigma, 1, \mu)$: The Lévy distribution: This distribution is concentrated on (μ, ∞) and has law
$$\left(\frac{\sigma}{2\pi}\right)^{1/2}\frac{1}{(x-\mu)^{3/2}}\exp\left\{-\frac{\sigma}{2(x-\mu)}\right\}\mathbb{I}_{x>\mu} \tag{3.85}$$

Even though closed form solutions for the probability densities are only known for these three specific cases, there are closed form algorithms (see Section 4.4) for simulating random variables on $\mathbb{R}$ regardless of the values of the parameters [25, 7]. One sided stable processes – processes with positive or negative jumps only – arise in the cases where $\beta = \pm 1$.

This class of processes was studied by Paul Lévy in the 1920's. Since then these distributions have been applied in economics, finance, communications systems and cosmology [71].

[5] and has nothing to do with the variance when $\alpha < 2$

[6] note this is not true when $\alpha = 1$

3.8 The Tempered Stable Process

The tempered stable process is an extension of stable processes through 'tempering' of the larger jumps. This happens by multiplication of the Lévy measure by means of a monotonically decreasing function - typically the exponential function. The behaviour of the small jumps remains more or less unchanged, while the bigger jumps are reduced in frequency when compared to the stable case. The tempered stable subordinator has Lévy measure and Laplace exponent given by [25]:

$$\nu(x) = \frac{ce^{-\lambda x}}{x^{\alpha+1}}\mathbb{I}_{x\geq 0} \tag{3.86}$$

$$\ell(u) = \begin{cases} c\Gamma(-\alpha)[(\lambda-u)^{\alpha}-\lambda^{\alpha}] & (\alpha \neq 0) \\ c\log(1-u/\lambda) & (\alpha = 0) \end{cases} \tag{3.87}$$

where c and λ are positive constants and $0 \leq \alpha < 1$. The case $\alpha = 0$ is simply the gamma process, and is included even though the gamma process cannot be obtained by exponential tilting of a stable process. By time changing an independent Brownian motion with a tempered stable subordinator, one obtains the *normal tempered stable process*. The density is generally not known, although it does admit a series representation. The characteristic function is given by:

$$\phi_{TS}(u;\kappa,a,b) = \exp(ab - a(b^{1/\kappa} - 2iu)^{\kappa}) \tag{3.88}$$

The distribution is infinitely divisible. Its series representation is given by[76]:

$$f_{TS}(x;\kappa,a,b) = \exp(ab - \frac{1}{2}b^{1/k}x)\frac{1}{2\pi a^{1/k}}\sum_{k=1}^{\infty}(-1)^{k-1}\sin(k\pi\kappa)\frac{\Gamma(k\kappa+1)}{k!}2^{k\kappa+1}\left(\frac{x}{a^{1/k}}\right)^{-k\kappa-1} \tag{3.89}$$

Both the gamma distribution ($\kappa \to 0$) and the inverse Gaussian ($\kappa = 0.5$) distribution are special cases of the TS process.

The class of tempered stable distributions were proposed by Tweedie [80] and have been used in survival analysis. They have also been studied by Hougaard [45] and Barndorff-Nielsen et. al. In general, stable distributions have always been overlooked because of the fact that their densities scarcely have a closed form solution. With technological advances and computational improvements, simulating these processes and approximating them numerically have become very reasonable options and they are worth looking into as models for financial markets.

3.9 The CGMY Process

This process is a generalisation of the variance-gamma process, which can be formulated in terms of three parameters C, G and M, by introducing a fourth parameter, Y. This parameter allows a whole class of infinite activity but finite variation Lévy processes to be created. The CGMY process was introduced by Carr *et al* [20] and can be extended to a six parameter case. The Lévy measure is given by:

$$\nu(dx) = \begin{cases} \frac{C\exp(Gx)}{|x|^{1+Y}}dx & (x<0) \\ \frac{C\exp(-Mx)}{x^{1+Y}}dx & (x>0) \end{cases} \tag{3.90}$$

The variance-gamma process is recovered for $Y = 0$, and this class of processes also serves as a generalisation of the model due to Kou, obtained when $Y = -1$. A closed form solution for the characteristic function can be computed, but the expression is somewhat complicated.

For the case where the process has finite variation, excluding the variance-gamma process, the characteristic function is given as:

$$\varphi(z) = (r - q + \mu)iz - \frac{\sigma^2}{2}z^2 + C\Gamma(-Y)\{(M - iz)^Y - M^Y + (G + iz)^Y - G^Y\} \tag{3.91}$$

where r is the risk-neutral drift, q is the dividend yield and μ is the real world drift of a particular stock. The C parameter is a measure of the overall activity of the process, while parameters G and M control for the skewness. Lastly, Y is a measure of the fine structure of the process. The easiest way to simulate the process is by means of a Poisson process approximation plus a Gaussian component to account for the small jumps (when $Y > 0$). The CGMY process is useful for simulating infinite activity but finite variation processes which differ from (are more general than) the variance-gamma process. The greater flexibility there is compensated for by the fact that they require considerably greater effort to simulate properly.

3.10 The Meixner Process

This process was introduced by Schoutens and Teugels [77] with Grigelionis [42] proposing its use for modelling stock returns. The density of the Meixner distribution $\mathcal{M}(\alpha, \beta, \delta)$ is given by:

$$f_{\mathcal{M}}(x; \alpha, \beta, \delta) = \frac{2\cos(\beta/2)^{2\delta}}{2\alpha\pi\Gamma(2\delta)} \exp\left(\frac{bx}{a}\right) \left|\Gamma\left(\delta + \frac{ix}{\alpha}\right)\right|^2 \tag{3.92}$$

The parameters have the following restrictions: $\alpha > 0, -\pi < \beta < \pi, \delta > 0$. The characteristic function looks as follows:

$$\phi_{\mathcal{M}} = \left(\frac{\cos(\beta/2)}{\cosh((\alpha u - i\beta)/2)}\right)^{2\delta} \tag{3.93}$$

The moments of the Meixner process lead to the following expressions for mean, variance, skewness and kurtosis [75]:

$$\mu = \alpha\delta\tan(\beta/2) \tag{3.94a}$$

$$\sigma^2 = \frac{\alpha^2\delta}{2}(\cos^{-2}(\beta/2)) \tag{3.94b}$$

$$\varsigma = \sin(\beta/2)\sqrt{2\delta^{-1}} \tag{3.94c}$$

$$\kappa = 3 + \frac{3 - 2\cos^2(\beta/2)}{\delta} \tag{3.94d}$$

The Meixner process is easily seen to be infinitely divisible since for any positive integer n,$\phi_{\mathcal{M}}$ is also the n^{th} power of a characteristic function (a Meixner process with parameters $(\alpha, \beta, \delta/n)$) [75]. Thus, the Meixner process can be associated with a Lévy process. It can be shown [42] that the Meixner process has no Brownian component, and a pure jump part governed by the Lévy measure:

$$\nu(dx) = \delta\frac{\exp(\beta x/\alpha)}{x\sinh(\pi x/\alpha)}dx \tag{3.95}$$

The first parameter in the Lévy triplet is:

$$\gamma = \alpha\delta\tan(\beta/2) - 2\delta\int_1^\infty \frac{\sinh(\beta x/\alpha)}{\sinh(\pi x/\alpha)}dx \tag{3.96}$$

The process is, however, not of finite variation since $\int_{-1}^{1}|x|\nu(dx) = \infty$. The Moment Generating Function for this process is continuously differentiable around zero, and so moments of all orders

exist. The Meixner process has semi-heavy tails making it suitable for asset modelling. It originates from the theory of orthogonal polynomials. The Meixner $(1, 2\zeta - \pi, \delta)$ distribution is the measure of the Meixner-Pollaczek polynomials [75]. It turns out that the monic variation of these polynomials are martingales for the Meixner process $(\alpha = 1, \zeta = (\beta + \pi)/2, \delta = 1)$. The Meixner distribution is a special case of the generalised-z distributions, an infinitely divisible family of processes defined through their characteristic functions. The generalised-z distributions and the generalised-hyperbolic distribution form non-intersecting sets.

The Meixner process is also related to an infinite sum of independent gamma processes. The process defined by (Biane, Pitman and Yor [14]):

$$C_t = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{\Gamma_{n,t}}{(n - \frac{1}{2})^2} \tag{3.97}$$

has a Laplace transform (because it is a subordinator):

$$\mathbb{E}[\exp(-uC_t)] = \left(\frac{1}{\cosh \sqrt{2u}} \right)^t \tag{3.98}$$

and a resulting time changed Brownian motion B_{C_t} follows a $\mathcal{M}(2, 0, t, 0)$ distribution [75]. Again, this process is difficult to simulate and generally happens through discretisation of the Lévy measure.

Chapter 4

Simulation from Statistical Distributions

In a financial modeling context, sampling from statistical distributions is almost surely unavoidable. This arises from the fact that much of the underlying mathematics is drawn from probability theory thus requiring a vast collection of quantitative techniques such as the simulation of random variables from specific types of distributions. In many cases this necessitates using specific algorithms and attempting to implement these methods on a computer effectively. In this, it is also hoped that the abstract and intricately complex models can be bridged with the real, everyday world of finance without losing too much in the translation – still making these models relevant enough to be used on a daily basis. The ability and efficiency of the algorithms to simulate from the desired statistical distributions is key to the overall strength of the model – as is so often said: 'a chain is only as strong as its weakest link'. Often, modellers, for the sake of convenience, choose to settle for something which 'works' and are more concerned with the end product or result and not necessarily how accurate the 'working' process has actually been. Such a naive 'black-box' style approach to modelling can be hazardous, and can lead to vastly spurious and incorrect results.

It is also important to be aware of the dimensionality of the problem at hand. This depends on how the problem and model in question are set up, but is affected by two quantities: the number of assets and number of time steps used in the model. It is not uncommon to observe financial engineering problems with a high dimensionality. The square root convergence of Monte Carlo, methods mentioned in Chapter 1, can only be improved on by use of superior sampling methods or, under special circumstances, Quasi-Monte Carlo methods. It may even be possible to reduce the dimensionality of a problem to the *effective dimensionality* by means of a principal components analysis and this is certainly an area for further investigation.

This chapter begins with a survey of uniform number generators and suggests that the Mersenne Twister is the algorithm of choice, since it is fast, can be used for high-dimensional problems and has a practically infinite period. Then, some attention is given to the theory and construction of low discrepancy sequences and how they can be used to reduce variance and improve convergence. Their application is not always fruitful, however. The issue of sampling from non-uniform distributions is then tackled; firstly by discussing various transformation techniques for uniform variates, and then by giving algorithms which have proven to be useful for the purposes of this dissertation. The theory of certain algorithms such as Marsaglia and Tsang's gamma generator and L'Ecuyer and Simard's symmetric beta generator are also discussed in detail.

4.1 Pseudo-random number generation

There is a crucial reliance on techniques involving the transformation of uniform random variables in simulating from different statistical distributions in a random fashion. Therefore, first and foremost, it is of utmost importance that reliable and accurate uniform sampling techniques are used. If a poor underlying uniform distribution sampling technique is used then this will result in poor transformed variates. Various different and well documented methods for generating uniform variates exist, and many software packages these days come equipped with the ability to generate *pseudo-random* numbers. However, anyone remotely serious about performing Monte Carlo simulations would be advised to spend some time investigating the possibility of including their own serious random number generator, as it is not necessarily given that software developers have done their homework and put in a decent uniform random number generator. Generating uniformly generated numbers requires a lot of careful attention and is remarkably simple to do badly, thus reliance on built in generators should be tentative at best.

The standard random number generator in Matlab from version 7.4 onwards, for example, is the Mersenne Twister, which falls into the category of pseudorandom number generators known in the literature as *generalised feedback shift register generators* (GFSR). Its authors, M. Matsumoto and T. Nishimura [65], claim that it exhibits equidistribution properties in 623 dimensions. This has important implications for Monte Carlo simulation in that the higher this embedding dimension, the safer the underlying number generator will be. In addition to reflecting the fact that this generator belongs to the 'twisted' GFSR class, its name also captures the property that its period is a Mersenne number (a prime number which can be written as $2^n - 1$, for some $n \in \mathbb{N}$). The period of the Mersenne Twister is $2^{19,937} - 1$; an awfully large number which practically can be assumed to be infinite. As Peter Jäckel puts it in [49]:

> *"In order to give you a feeling for this number, imagine that we started at the time of the creation of the universe a computer producing 1 billion numbers per second from the Mersenne twister sequence. The fraction of the full period that this computer would have produced by now is a decimal number with 5975 digits of zeros behind the decimal point, prior to any non-zero digits. In other words, this computer could continue to draw numbers for many thousand lifecycles of your average solar system between its formation and collapse into a black hole before beginning again."*

The default in earlier versions of Matlab (from version 6 onwards) is to use a method due to George Marsaglia, a generator belonging to the class "subtract-with-borrow". The simplest form of all these generators being mentioned here is the Linear Congruential Generator. In these generators, each single number determines its successor. The form of the generator [39] is (with $0 \leq x_i < m$):

$$x_i \triangleq (ax_{i-1} + c) \pmod{m} \tag{4.1}$$

Here m is known as the modulus, a is the multiplier and c is the increment. Here, a and c do not need to be less than m, since it is always possible to find $a', c' < m$ such that $ax + c \triangleq a'x' + c' \bmod m$). The pseudo-random number in this case is given by $\frac{x_i}{m}$. A few properties of a LCG are that, firstly, they can have at most period m. This will only happen if c and m are relatively prime, for example. The ability of LCG's to produce decent pseudo-random numbers depends critically on the choice of c, m and a. For Monte Carlo simulations, these types of generators are not recommended since the successive values exhibit serial correlation, which can easily be tested in a program such as Matlab or MSExcel. Thus for the purposes of this dissertation they are not considered further, other than to note that they are the building blocks for most other pseudo-random generators.

When a pseudo-random generator is used, it is a straightforward task to show that in the unit hyperplane (Figure 4.1) (even for low dimensions) the phenomenon of 'clustering' is observed. This crudely means that there appear to be certain 'holes' that will never be filled, and certain points around which more numbers seem to be generated than others.

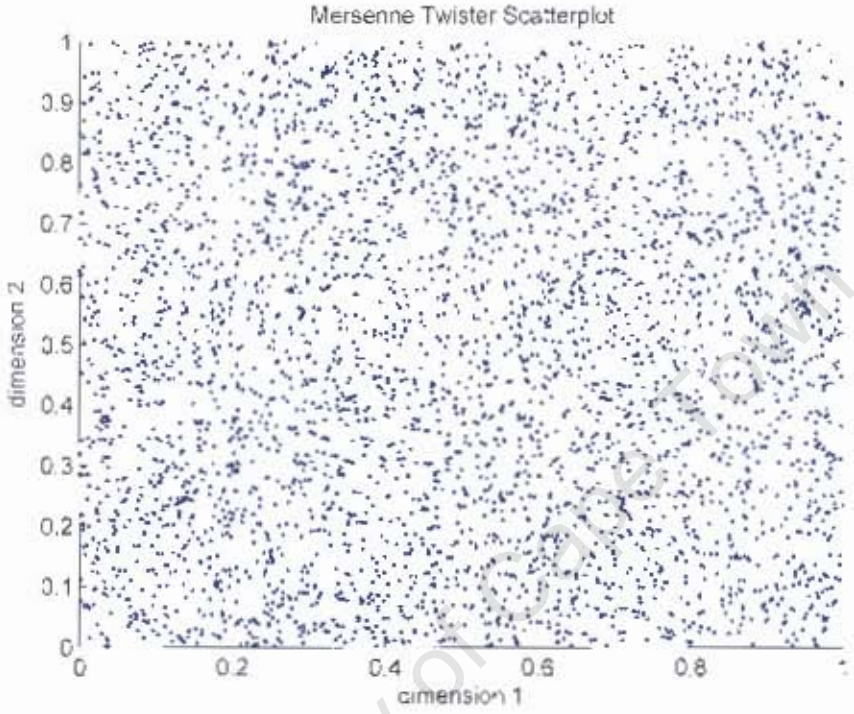

FIGURE 4.1: Scatter plot of 4000 pairs of pseudo-random numbers generated using the Mersenne Twister

4.2 Low discrepancy number sequences

Number sequences which are said to be *low discrepancy*, do not aim to be serially uncorrelated, like their pseudo-random counterparts. One of the reasons for their taking previous vector draws into account is so as to avoid the inevitable 'clusters' which are commonly observed by use of a(ny) pseudo-random number generator. The theory of these sequences can be dated back to H. Weyl in 1917 [49], and since then many such sequences have been studied and have found popularity and usage in Quasi-Monte Carlo techniques. These sequences play an important role in cases where closed form solutions of integrals are not available, which would allow for the application of variance reduction techniques. The discrepancy of a sequence of numbers is a measure of how inhomogeneously a set of $d-$dimensional vectors is distributed in the unit hypercube [49]. A simple geometric interpretation is given as follows: generate N pairs of multivariate draws, $\mathbf{r}(i)$, using a selected method of uniform number generation. Then, select a sub-hypercube, $S(\mathbf{y})$, which has a hyper-rectangular domain from $\mathbf{0}$ to $\mathbf{y}$ and let $n_{S(\mathbf{y})}$ denote

the number of draws from N which are in $S(\mathbf{y})$. This is written, mathematically as:

$$\begin{aligned} n_{S(\mathbf{y})} &= \sum_{i=1}^{N} \mathbb{I}_{r_i \in S(\mathbf{y})} \\ &= \sum_{i=1}^{N} \prod_{k=1}^{d} \mathbb{I}_{y_k \geq r_{ik}} \end{aligned}$$

In the limit as $N \to \infty$, a definite requirement of the generator is perfect homogeneity, meaning that for a perfectly homogeneous and uniform distribution on a unit hypercube, the probability of being in a subdomain is equal to the volume of that subdomain.

$$\lim_{N \to \infty} \frac{n_{S(\mathbf{y})}}{N} = \prod_{i=1}^{d} y_i$$

So, for a given $\mathbf{y}$, one can compare $n_{S(\mathbf{y})}/N$ with the volume of $S(\mathbf{y})$, which leads to a quantification of this notion of discrepancy. Using the L_2-norm, one obtains

$$T_N^{(d)} = \left(\int_{[0,1]^d} \left(\frac{n_{S(\mathbf{y})}}{N} - \prod_{k=1}^{d} y_k \right)^2 dy \right)^{\frac{1}{2}} \tag{4.2a}$$

Similarly using the $L_\infty-$norm, another discrepancy measure can be defined:

$$D_N^{(d)} = \sup_{y \in [0,1]^d} \left| \frac{n_{S(\mathbf{y})}}{N} - \prod_{k=1}^{d} y_k \right| \tag{4.2b}$$

By the nature of the underlying norms involved, one can state that $D_N^{(d)} \geq T_N^{(d)}$, but in practice evaluation of the $L_\infty-$norm is tedious. Here, a number-theoretical definition of a low-discrepancy sequence is given:

Definition 4.1 (Low-discrepancy sequence)**.** A sequence in $[0,1]^d$ is said to be a *low-discrepancy* sequence if for all $N > 1$ the first N points in the sequence satisfy

$$D_N^{(d)} \leq c(d) \frac{(\ln N)^d}{N} \tag{4.3}$$

for some constant $c(d)$ which depends only on d [49].

The idea here is that low discrepancy sequences have the points on a unit hypercube spaced and stretched out as far as possible such that they are equidistributed. They aim to cover this unit hyperplane consistently, fairly and evenly. In doing this they become serially correlated. A discussion on various types of low discrepancy numbers is now presented. Good reading sources for these sequences are Jäckel [49] and Glasserman [40].

4.2.1 Halton Numbers

The Halton number sequence is simply a higher dimensional extension of a sequence known as the *van der Corput* sequence, introduced in 1935 by J. G. van der Corput. It was shown by Neiderreiter [70] (Theorem 3.6) which proves that all such sequences are low-discrepancy sequences. It is constructed by simply reversing the base n representation of the sequence of natural numbers. For example the binary van der Corput sequence starts as 0.1_2, 0.01_2, 0.11_2, 0.001_2, 0.101_2, ...,

which are simply the reversed binary forms of $1, 2, 3, 4, 5, \ldots$. Although one can technically do this for any base n, a prime integer base can be chosen so as to reduce pairwise asymptotic periodicity. When (possibly more than one) prime base is used, the resulting sequence is called a Halton sequence.

For a $d-$dimensional Halton sequence, d prime numbers must be chosen and some initial seed of a sequence, $\gamma_0, \in \mathbb{N}$ must be specified. Then each draw of the Halton sequence is calculated as follows:

- For that particular element of the sequence, $\gamma(n)$, calculate the representation of that integer in each of the prime bases. In other words find the coefficients a_k in the associated prime base:
$$\gamma(n) = \sum_{k=1}^{m_{ni}} a_{ki} p_i^{k-1} \tag{4.4a}$$
- Invert the sequence of calculated coefficients, and use these as multipliers of fractions in the prime number base. This then yields the Halton number for that particular prime base, for the draw $\gamma(n)$:
$$u_{ni} = \sum_{k=1}^{m_{ni}} a_{ki} p_i^{-k} \tag{4.4b}$$

Example 4.1. The following example may help to make things slightly clearer. Suppose we wish to find the Halton number corresponding to the integer 63 using the prime number 5 as the base. In this base, 63 has the following representation:

$$\begin{aligned} 63 &= 2.5^2 + 2.5^1 + 3.5^0 \\ &= 223_5 \end{aligned}$$

Thus to calculate the corresponding Halton number, invert 223_5 to obtain 0.322_5, and convert this to a decimal number:

$$\begin{aligned} 0.322_5 &= \frac{3}{5} + \frac{2}{25} + \frac{2}{125} \\ &= 0.6960_{10} \end{aligned}$$

Thus the corresponding Halton number is 0.6960. Use of a different prime base will produce a totally different Halton number. To generate a $d-$dimensional pseudo-random vector, the same expansion of one particular integer is carried out to d different prime bases. The Halton number associated with 63 expanded to base 7 is 0.21. Thus, a 2-dimensional Halton number draw could be associated with prime bases 5 and 7, and the pseudo-random vector would then be [0.696, 0.21]. Generally, consecutive prime bases are used, so a $d-$dimensional Halton number draw would use the first d prime numbers.

Halton numbers work relatively well for low dimensions, but they fail to cover the unit hypercube at all well for high dimensions. Part of the reason is that in using higher dimensionality, the space (or volume, if you like) required to be filled becomes very large. Also, for bigger prime bases, the period of the associated Halton sequences become large and thus more iterations are required to cover the unit hypercube. This can be seen graphically in Figure 4.3 Finally, the correlation of these sequences becomes unacceptably high, for high dimensions as can easily be verified by a simple scatterplot of two of the higher dimensions.

Various methods have been proposed to deal with this problem, of which the most famous is

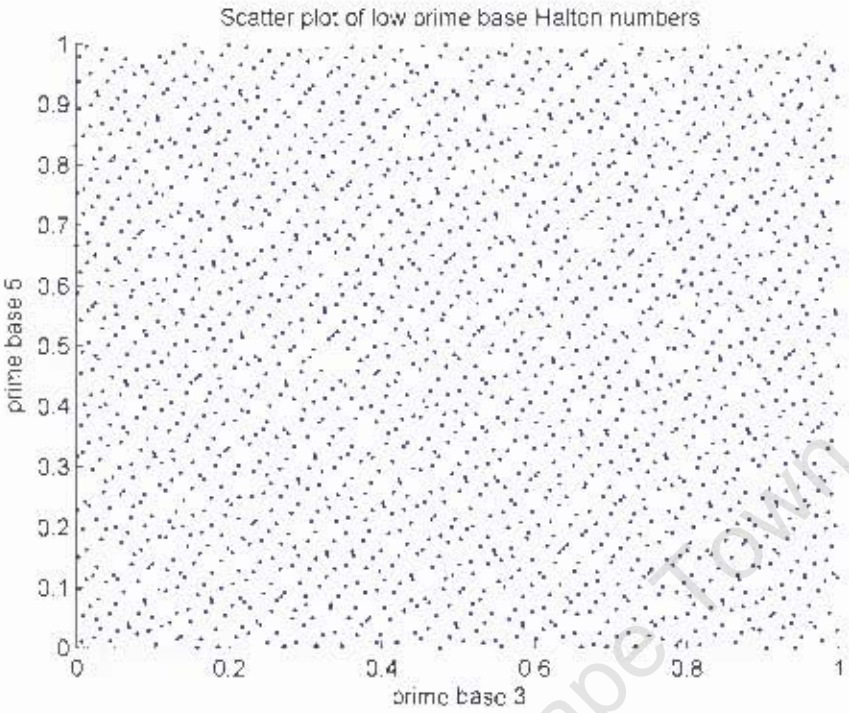

FIGURE 4.2: Scatter plot of 2 dimensional low prime base Halton sequence

probably the scrambled Halton sequence – using special predetermined permutations (see Faure sequences) of the coefficients obtained in constructing the standard sequence. Another option is to use randomly shuffled versions of the standard 1-dimensional sequence in constructing higher dimensional sequences. Also, for larger prime number bases, reasonably large starting seeds should be used to assist in avoiding the introduction of unnecessary correlations. Finally, a *leaped Halton sequence* can be constructed which multiplies each term in the input integer sequence or *Grey code* by a constant number which is relatively prime to all the prime integer bases being used.

4.2.2 Faure Sequences

These are similar to the Halton sequences in many ways, but they only make use of one base for all dimensions and they use permutations of the vector elements for each dimension. The base of a Faure sequence is the smallest prime, larger than or equal to the dimensionality of the problem: for a 50 dimensional problem the Faure sequence will use the prime number 53, while the Halton sequence will use the first 50 prime numbers. The 50^{th} prime number is 229. Faure sequences suffer a similar problem of slow generation speed on an increasingly finer grid of the unit hypercube with an increase in dimension although the problem is slightly ameliorated because Faure sequences always use smaller prime numbers.

Reordering the sequence within each dimension helps to reduce some of the correlation problems induced by high dimensionality. The algorithm uses the same style equations given by (4.4), but before (4.4b) can be used, the coefficients are rearranged or permuted using a recursive equation:

$$a_i^d(n) = \sum_{j\geq 1}^{m} \binom{i}{j} a_j^{d-1}(n) \pmod{b} \tag{4.5}$$

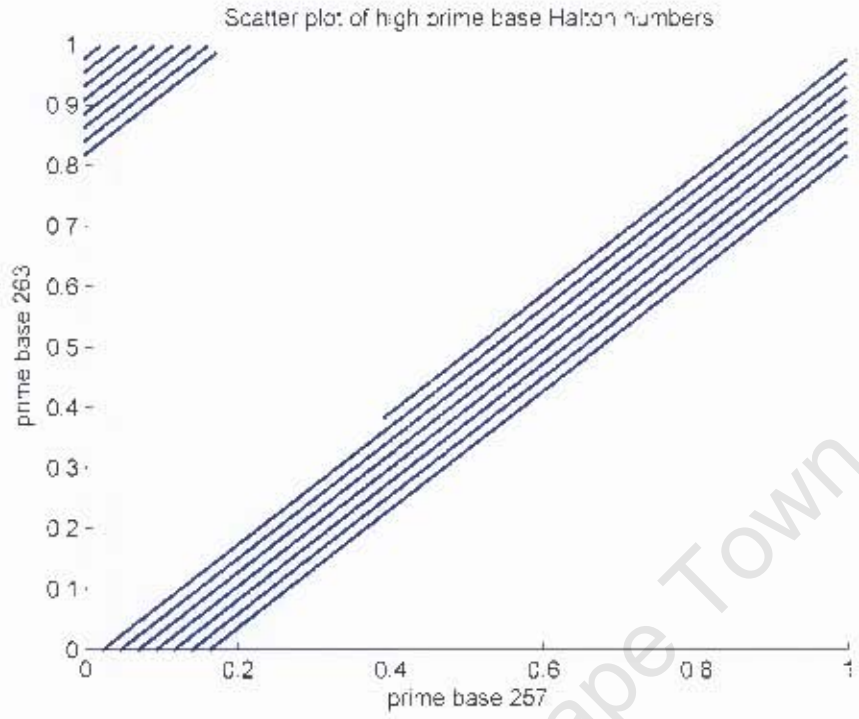

FIGURE 4.3: Scatter plot of 2 dimensional high prime base Halton sequence

To use the above recursive formula, generate the first dimension using a basic *van der Corput* sequence with the same base as the desired Faure sequence, and then reorder the numbers using the above equation for the higher dimensions. Jang and Galanti [50] report that Faure sequences can exhibit clustering around zero, and to avoid this it is suggested that the first $(b^2 - 1)$ points are dropped. Needless to say, the Faure sequences do tend to display the same unpleasant correlation characteristics for high dimensions, making their use less desirable.

Faure himself introduced these sequences [34] and showed that they were low-discrepancy sequences. The constant $c(d)$ in (4.3) can be shown to only depend on d and not on n aswell and it tends to zero very quickly in contrast to Halton sequences [34, 40].

4.2.3 Sobol' Sequences

These sequences were introduced by I. Sobol' in 1976 [78]. Although they are similar to the previous two types of number sequences, the Sobol' numbers have a slightly more involved construction, and careful attention is required in the programming to get it right. Like the Faure and Halton sequences the Sobol' sequences are based on the Van der Corput sequence. The difference is that the Sobol' sequences are constructed exclusively in base 2. Faure sequences are $(0, d)$ sequences in a base at least as large as d. The first dimensional Sobol' sequence is thus identical to the *van der Corput* sequence in base 2. The higher dimensional sequences again are permutations of the sequence of the first dimension, and as a result of the shorter cycle length due to the use of base 2 there is a computational time advantage and the arithmetic is efficient computationally.

The permutations themselves depend on a set of direction numbers: $v_b = \frac{m_b}{2^b}$. The values for m_i are chosen to satisfy a recurrence relation using the coefficients of a primitive polynomial in

the Galois field (GF) of order two [66]. As a starting point for these polynomials, a definition is required.

Definition 4.2. Two integers i and j are said to be congruent with respect to the modulus m, iff the difference $i - j$ is divisible by m. This is expressed as:

$$i \triangleq j \pmod{m}$$

A polynomial $P(z)$ of degree g is considered an element of the ring $GF[m, z]$ of polynomials over the finite field $GF[m]$, if it is assumed that all the coefficients $a_k \in GF[m]$.

$$P(z) = \sum_{j=0}^{g} a_k z^{g-j}$$

This means that all algebra on the coefficients a_k must be carried out modulo m. Such a polynomial is considered to be *irreducible* modulo m if there are no two other (non-constant) polynomials $Q(z)$ and $R(z)$ which are not equal to $P(z)$ such that:

$$P(z) \triangleq Q(z)R(z)$$

An irreducible polynomial modulo m in $GF[m, z]$ is the equivalent of a prime number in the set of integers. The order of a polynomial $P(z)$ modulo m is given by the smallest positive integer q for which $P(z)$ divides $z^q - 1$. An irreducible polynomial of degree g is considered to be *primitive* modulo m if it has order $m^g - 1$.

For each of the dimensions d the basis of number generation is given by *direction numbers* of which there is one for each of the b bits. Associated with each dimension $k \in \{1, \ldots, d\}$ is a primitive polynomial P_d and the first g direction numbers[1] can be chosen freely. Once these have been chosen the remaining direction numbers from $g + 1$ to b are chosen by means of a recurrence relation defined from (4.2.3). These are determined as follows:

$$\nu_{kl} = \frac{\nu_{k(l-g)}}{2^g} \oplus_2 \sum_{j=1}^{g}{}^{\oplus_2} a_{kj}\nu_{k(l-j)} \tag{4.6}$$

where $\oplus_2$ denotes XOR (exclusive or) addition modulo 2 without carrying. Note that $\sum^{\oplus_2}$ also denotes a whole series of XOR additions. In a nutshell constructing a Sobol' sequence consists of the following steps:

- Choose a starting integer γ randomly, as the starting seed of the sequence;
- Calculate the *Grey Code* of γ. This is defined as $G(\gamma) = \gamma \ominus_2 [\gamma/2]$. Then convert $G(\gamma)$ to its binary representation;
- Perform a bitwise XOR sum of the direction numbers associated with the digits of $G(\gamma)$ which are different from zero (counting from right to left). For example if digits 1 and 4 in $G(x)$ are different from zero, then an XOR sum must be done on the first and fourth direction numbers;
- Convert this number back to a decimal number - this is the first Sobol' number in that particular dimension.

[1] note that g does depend on k

The above algorithm can be improved upon to reduce computational time (Antonov & Saleev 1979 [6]) yet, while Sobol' number generation is computationally quicker than the Halton or Faure methods, it is not without its problems. From a correlation point of view, Jang and Galanti [50] showed that Sobol' sequences preserve their uniformity properties reasonably well up to a dimension of 260, but in these high-dimensional states the Sobol' numbers tend to repeat themselves across dimensions leading to clusterings. One way to deal with this is to discard the first n points where n is some arbitrary integer (Boyle, Broadie & Glasserman in a paper written in 1995 suggest $n = 64$).

Using a routine written in Matlab [12], 2000 Sobol' numbers in 16 dimensions (so 32000 numbers in total) were generated in a time of 1.594 seconds on a Intel Centrino 1.5GHz Pentium M processor. Some scatter plots for different dimensions are shown.

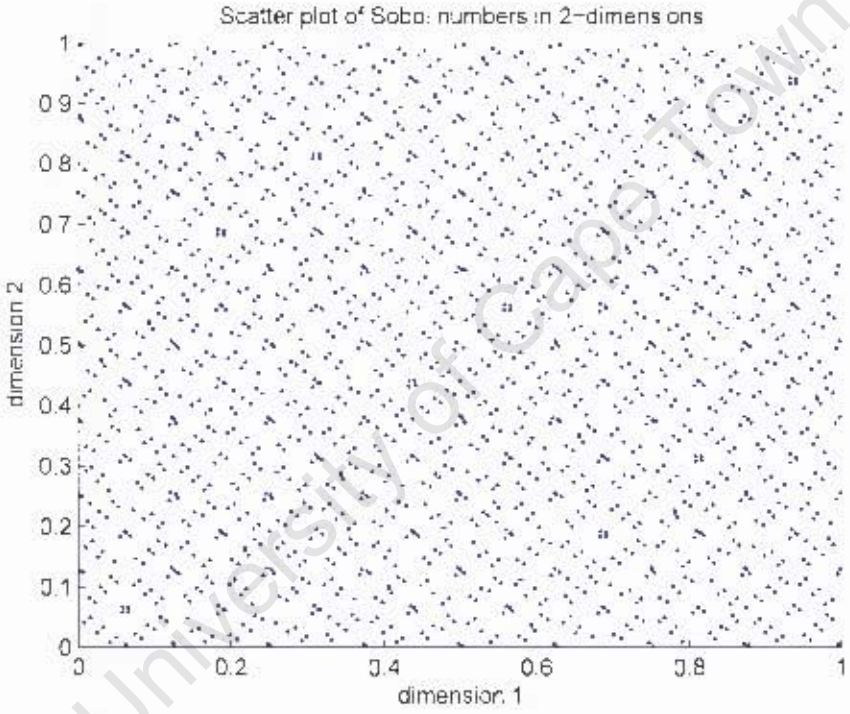

FIGURE 4.4: Two dimensional scatterplot of a low dimensional Sobol' sequence

4.3 Extending sampling beyond the uniform distribution

Once one has a means of obtaining a sequence of uniform random variables, it is then possible to simulate random variables which follow specific (and possibly arcane) distributions. The three major methods used are:

1. Inverse Transform Method
2. Acceptance / Rejection techniques
3. Ratio of uniforms method

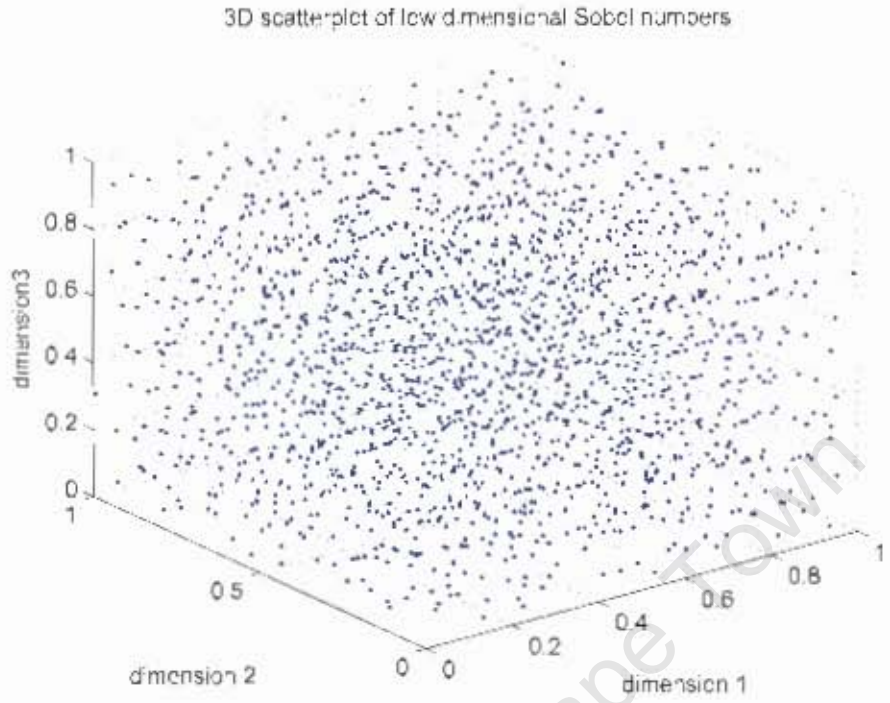

FIGURE 4.5: Three dimensional scatterplot of a low dimensional Sobol' sequence

4.3.1 The Inverse Transform Method

For a particular probability distribution, the *cumulative distribution function* $F(x)$ is defined as:

$$F_X(x) = \mathbb{P}[X \leq x] \tag{4.7}$$

When the distribution has a well defined *probability density function*, $f(x)$ (which is not always the case), (4.7) can also be written as

$$F_X(x) = \int_{\infty}^{x} f(x)dx \tag{4.8}$$

The integral over the domain of such *probability density functions* is equal to 1. Also, $\text{ran}(F) \subset [0,1]$ (i.e $F : [a,b] \to [0,1]$, with $a, b \in \mathbb{R} \cup \pm\infty$). If $F(x)$ is strictly increasing (so $F(x+h) > F(x) \quad (\forall h > 0)$) and continuous, then it is one-to-one and it is possible to find the well defined inverse $F^{-1}(x)$. Under these assumptions, $F^{-1} : [0,1] \to [a,b]$ and can be shown to be a uniform random variable taking on values in the unit interval [39]. If the inverse function has a closed form solution and one has access to some form of uniform random number generator, one can use the inverse map to generate variables which follow the desired law of $\mathbb{P}_X$. For cases where $F(x)$ is not strictly increasing, or where it is not one-to-one, a little care has to be taken when defining the inverse function, but it is still possible to do so.

By design, the *inverse transform method* is simplistic and accurate, but it does not always work - motivating the use of alternate methods mentioned above. An easy enough example illustrating this is given by the normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{4.9}$$

with CDF,

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \tag{4.10}$$

Unfortunately, integrals of this type have no closed form solution. The integral representation is generally known as the *error function* which has to be evaluated numerically. Efficient numerical algorithms exist (see for example work by Peter Acklam [3], or Numerical Recipes [1]) for calculating the error function, but the point is exactly that – numerical integration (or a similar technique) is required. This means that the inverse function can not be readily defined, since the inverse error function is not going to magically have a closed form solution.

Today, even though many distributions do not permit a closed form representation of their CDF's, it is still possible to evaluate them numerically with remarkable efficiency[49]. This is largely thanks to the efficiency of computers and the modern microprocessor, and explains why sampling from the normal distribution these days is quick and effective.

In general, the process works as follows: for a fixed set of distribution parameters, various points of the CDF can be calculated, and an interpolation table can be set up using one of many splining techniques (cubic, monotonicity preserving etc.). The method of sampling from this particular distribution would be as follows:

- Draw a uniform random variable, u
- Use (cubic) spline interpolation to get an estimate for F^{-1}
- Use Newton-Rhapson or (even better - faster convergence) Halley's method to solve $F(x) - u = 0$

Jäckel [49] suggests that this method should be preferred and Gentle [39] suggests that it may be better than any other method. It is particularly useful if a large number of variates are required for a distribution with fixed parameters. However, if the parameters change from call to call (which does seem to be the case for most scenarios involving Lévy process) setting up the required interpolation table each time and other overhead costs may make this method unfavourable.

4.3.2 Acceptance / Rejection method

These methods make use of a sampler density when it is difficult or impossible to calculate the inverse CDF. This has admittedly become less of a problem in recent years. Instead of drawing directly from the desired target density $\psi(x)$, one samples from a hopefully similar density $\tilde{\psi}(x)$, from which variates can be easily obtained. Simulation is then carried out using variates from the sampler density, and each function evaluation is corrected according to the likelihood ratio of the target density and the sampler density [49]. This, mathematically, is equivalent to saying the following:

$$\int f(x)\psi(x)dx = \int f(x)\left(\frac{\psi(x)}{\tilde{\psi}(x)}\right)\tilde{\psi}(x)dx \tag{4.11}$$

For this method, a sampler density, $\tilde{\psi}(x)$, is required such that there is a scalar multiple of it which dominates $\psi(x)$. That is, there exists a c such that $c\tilde{\psi}(x) \geq \psi(x)$ for all x in the domain of ψ. The essential idea of rejection sampling is then as follows [39, 49]:

- Draw a variate x from the sampling density, $\tilde{\psi}(x)$
- Compute the value of both $\tilde{\psi}(x)$ and $\psi(x)$ at x

- Draw a uniform variate $u \sim \mathcal{U}(0,1)$
- If $u \cdot c\tilde{\psi}(x) > \psi(x)$ then reject this attempt and start again.

Obviously, the speed and efficiency of the algorithm depends on how close $\tilde{\psi}(x)$ is to $\psi(x)$. It is also possible to reduce the number of computations in acceptance / rejection testing. This generally is achieved through the use of a 'squeeze' function [62]. Most algorithms which use a squeeze function usually only use one below the density of interest, allowing for quicker rejection [39]. There are also variations on the basic method described above, including the *transformed acceptance / rejection method* [82] and Marsaglia's *exact-approximation method* [61]. These sample from a density proportional to $p(H(x))h(x)$, where $H(x) = G^{-1}(x)$, with G the CDF corresponding to the sampling density, and $h(x)$ is the derivative of H with respect to x. These algorithms then return the value of $H(x)$. The choice of H for the *exact-approximation method* is crucial to the efficiency of the algorithm: the closer it is to the inverse of the CDF of the target distribution, the better. These algorithms can quite readily be extended to multivarite situations, which is not the case with the inverse CDF method [39]. However, these algorithms are not without their issues: there is an implosion of the sampling yield with an increase in dimensionality and it is difficult to amend them to efficient application of low-discrepancy numbers [49].

4.3.3 Ratio of Uniforms method

A useful relationship between random variables U, V and U/V is described in Kinderman and Monahan [55]. If (U, V) is uniformly distributed over the set

$$C = \left\{(u, v); 0 \leq u \leq \sqrt{h\left(\frac{v}{u}\right)}\right\} \tag{4.12}$$

where h is a non negative integrable function, then V/U has probability density proportional to h. In practice a rectangle R may be chosen which encloses C, a point in the rectangle may be generated and then rejected if it does not satisfy

$$u \leq \sqrt{h\left(\frac{v}{u}\right)} \tag{4.13}$$

The method is simple and easy to apply. It has been used to generate gamma, Poisson, binomial and student-t random variables, to name but a few [39]. It is possible to generalise the method to certain multivariate distributions. In certain instances, the quality of the pseudo-random output stream from a ratio of uniforms method can be poor. This has to do with the fact that the method transforms all points which lie on one line through the origin into a single number. The lattice structure from linear congruential generators, for example, means that these lines passing through the origin have regular patterns resulting in structural gaps in the output from the ratio-of-uniforms method. Hörmann and Derflinger [44] have studied this issue and based on their analysis, they recommend the transformed rejection method over the ratio of uniforms method. The quality of the ratio of uniforms method, however, is more a function of the quality of the uniform generator used.

4.4 Simulating from specific distributions

Thus, having looked at generating uniform numbers along with methods of transforming them to other distributions, attention is now given to specific sampling algorithms for particular

distributions. These are important since they determine the overall accuracy of the numbers which will ultimately be used in simulating Lévy process paths. Use of poor algorithms at this point, will reflect in the construction of associated Lévy processes.

4.4.1 Sampling from the normal distribution

One of the most frequently used distributions across the statistical sciences is the Gaussian distribution, which is completely parameterised by two parameters: the mean and variance. The main challenge is to generate variates which follow the *standard* normal distribution, since these can be arbitrarily transformed to normal variates with any mean and variance. There are a wide variety of sampling techniques available, of which the most well known is probably the Box-Muller algorithm. More accurate polynomial approximations have been developed [3] for the cumulative normal inverse which are double precision.

Algorithm 4.1. *The Box-Muller method [17]:*

If s and t are independently distributed as uniform (0,1) and:

$$x = \sqrt{-2\log(s)}\cos(2\pi t), \tag{4.14}$$

$$y = \sqrt{-2\log(s)}\sin(2\pi t) \tag{4.15}$$

Then the joint distribution for x and y is given by [49]

$$\psi(x,y) = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \left(\frac{e^{-0.5x^2}}{\sqrt{2\pi}}\right)\left(\frac{e^{-0.5y^2}}{\sqrt{2\pi}}\right) \tag{4.16}$$

In other words, it is the distribution of two independent standard normal random variables. By noting that the trigonometric terms on the right hand side of the above expression are the abscissa and ordinate of a point on the perimeter of a unit circle one can construct an alternative method to the Box-Muller method above: draw a random point from within a unit circle, and then use its cartesian co-ordinates (s,t) as follows:

- Set $u = s^2 + t^2$. Check that $u \leq 1$, then set
- $x = s\sqrt{-2\frac{\ln u}{u}}$
- $y = t\sqrt{-2\frac{\ln u}{u}}$

To ensure that a cartesian coordinate pair is chosen so that they describe a point inside the unit circle, the uniform draw must take place on the interval (-1,1) = 2*(0,1) - 1, and then simply check the condition $s^2 + t^2 = u \leq 1$. If this condition holds true then there exists a θ such that $s = \sqrt{u}\cos\theta$ and $t = \sqrt{u}\sin\theta$. It is possible to show [12] that u is uniformly distributed on [0,1] and that $\theta = \arctan\frac{s}{t}$ is uniformly distributed on $[0, 2\pi]$.
Since the area of the unit circle is π and the area of a 2×2 square is 4, the rejection yield is $\frac{\pi}{4} \simeq 0.786$ [49]. The loop will succeed on the first try nearly 80% of the time. This method is often known as the *Polar-Marsaglia method*

Jäckel [49] mentions that rejection methods are quite dangerous, and should not be used in conjunction with low-discrepancy sequences (Section 4.2). He provides a graphical example for how the Box-Muller method can work quite well when the underlying uniform generator is the Mersenne Twister, but when Sobol' numbers are used, the resulting scatterplot is disastrous. A second problem with such highly sophisticated deterministic methods is that the interaction of two (or more) non-linear systems may undesirable interactions that are hard-to-foresee and difficult-to-imagine. This is known as the Neave effect, after being discovered in 1973 by H. R. Neave.

Acklam's method for calculating the inverse of the standard normal cumulative distribution function

Peter Acklam [3] has developed a computational algorithm for calculating the value of the cumulative normal inverse function, for a specific probability p. The output from the algorithm x has a relative error (compared to the true quantile) of less than 1.15×10^{-9} in the region where $x \geq -38$. Practically, this is feasible since the true probability of the cumulative normal density function when $x = -38$ is somewhere in the region of 2.88×10^{-15}. (The smallest number which can be represented by IEEE double precision arithmetic to full precision is 2.225×10^{-308}.)

His approximation breaks the domain up into three regions: a central region, and one for each of the tails. In each region a different rational minimax approximations is used. The central region is defined to be that for which $0.02425 < p < 0.97575$. One can use these rational approximations to produce an estimate of the desired (true) quantile x, which can then be refined using Halley's method (a third order root finding method, which is an extension of Newton's method and provides faster convergence).

For Lévy processes, random variates, from distributions other than the normal one, are required. The next few sections consider sampling algorithms from more general distributions.

4.4.2 Sampling from the Poisson distribution

Algorithm 4.2. *Simulating from a Poisson distribution with parameter θ, making use of the relation $(k+1)\mathbb{P}[N = k+1] = \mathbb{P}[N = k]\theta$ [25]*

- Set $p = \exp(-\theta), F = p, N = 0$
- while $U > F$, generate $U \sim$ Unif[0,1]
  1. $N \rightarrow N + 1$
  2. $p \rightarrow \frac{p\theta}{N}$
  3. $F \rightarrow F + p$
- RETURN N

Since the Poisson distribution is discrete, this algorithm calculates the value of the CDF at a particular point by summation of the probabilities up to a particular point, distributed uniformly on [0,1] as per definition of the CDF. A running counter is required and value returned by this counter is the quantity which follows a Poisson distribution.

4.4.3 Sampling from the gamma distribution

The gamma distribution possesses a very useful scaling property: if $S_t \sim \mathcal{G}(c, \lambda)$, then $\lambda S_t \sim \mathcal{G}(c, 1)$. Thus, it suffices to generate gamma distributed random variables for the case where $\beta = 1$ for any arbitrary value of α. When these parameters are put into the gamma probability density function (3.3) it simplifies down to

$$p(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x} \tag{4.17}$$

The above-mentioned scaling property can then be used. Typically, two types of cases are considered. one where $\alpha < 1$ and one where $\alpha > 1$. These generally require different algorithms which will be mentioned below. For the case where $\alpha = 1$, the distribution in question is simply

exponential, which has a closed form inverse cumulative distribution function. Sampling in this instance through the use of the inverse transform method is straightforward. One of the most promising methods found in this thesis, both quick and accurate, is described below. Although it is a rejection algorithm, it has a remarkably fast acceptance rate (which can be improved by making use of a 'squeeze' function). It turns out that this is the standard sampling method implemented in current versions of Matlab (7.0 and higher), which was the main environment used for simulations and calculations. The function name in Matlab is GAMRND.

Method of G. Marsaglia and W. Tsang (2001) for simulating gamma variates

Marsaglia and Tsang (2001) [63] provide an algorithm for generating Gamma variates as the cube of a suitably scaled normal random variable. Therefore, at the heart of this method lie two crucial requirements: a good uniform random number generator and a good Gaussian random variable generator.

The basic idea is to simulate random variables which follow a density $y^{\alpha-1}e^{-y}/\Gamma(\alpha)$. To this end, let:

$$h(x) = d(1+cx)^3 \tag{4.18}$$

for $-c^{-1} < x < \infty$. Here $d = \alpha - 1/3$ and $c = 1/\sqrt{9d}$. If random variable X is generated with density $h(x)^{\alpha-1}e^{-h(x)}h'(x)/\Gamma(\alpha)$ then $y = h(X)$ will have density $y^{\alpha-1}e^{-y}/\Gamma(\alpha)$ (This is known as Marsaglia's exact-approximation method, and the function used here involves the third power of a normal variate.) The goal is to find the representation of the function $h(x)$. A simple rejection method is being dealt with here, which makes the pesky normalisation constant $\Gamma(\alpha)$, disappear. Simplifying $h(x)^{\alpha-1}e^{-h(x)}h'(x)/\Gamma(\alpha)$ and rewriting in exponential form leads to the observation that a random variable, whose density is a normalising constant times $e^{g(x)}$, must be generated where:

$$g(x) = d\ln((1+cx)^3) - d(1+cx)^3 + d \tag{4.19}$$

The extra d is inserted in to g to make $e^{g(0)} = 1$. This simplification comes about as follows (constant terms are ignored / dropped):

$$\begin{aligned}
h(x)^{\alpha-1}e^{-h(x)}h'(x) &= \exp\left[\ln(h(x)^{\alpha-1} - h(x) + \ln(h'(x))\right] \\
&= \exp\left[(\alpha-1)\ln(d(1+cx)^3) - d(1+cx)^3 + \ln(3cd(1+cx)^2)\right] \\
&= \exp\left[(\alpha-\frac{1}{3})\ln((1+cx)^3) - \frac{2}{3}\ln((1+cx)^3) - d(1+cx)^3 + \ln((1+cx)^2)\right] \\
&= \exp\left[(\alpha-\frac{1}{3})\ln((1+cx)^3) - 2\ln((1+cx)) - d(1+cx)^3 + 2\ln((1+cx))\right]
\end{aligned}$$

Equation (4.19) then follows quite simply. The justification for this particular choice of function is further motivated by the following property:

$$e^{g(x)} \leq e^{-0.5x^2}, \quad -1/c < x < \infty \tag{4.20}$$

That is, g can be put under the unscaled normal density. It also turns out that this function is quite close to the normal distribution, occupying some 95.2% of it at $\alpha = 1$ to 99.7% of it at $\alpha = 10$ [63]. Another feature for this particular choice of function is that:

$$Ue^{-0.5x^2} < e^{g(x)} \iff \ln(U) - 0.5x^2 < g(x)$$

This avoids the use of the exponential function in the rejection method and gives a speed improvement, because calls to the natural logarithm function are faster than calls to the exponential

function. For the function $h(x)$ particular values of d, c and k now need to be chosen such that the density given above will be nearly normal. Ignoring normalising constants, the following can be done: Choose c, d and k such that

$$f(x) = e^{g(x)} = e^{(k\alpha-1)\ln(1+cx)-d(1+cx)^k+d} = e^{-0.5x^2+a_3x^3+a_4x^4} \tag{4.21}$$

for small values of a_3 and a_4. This is possible by choosing $d = \alpha - 1/k$, $c = 1/\sqrt{k^2\alpha - k}$. With these f becomes

$$f(x) = \exp\left[-0.5x^2 - \frac{(k-3)c}{6}x^3 - \frac{(k^2-6k+11)c^2}{24}x^4\right]$$

Evidently, $k = 3$. Then $d = \alpha - 1/3$ and $c = 1/\sqrt{9\alpha - 3}$. It is possible to choose $k = 4$ or even $k = 8$, making computation of $(1 + cx)^k$ easy. The algorithm outlined below makes use of the choice $k = 3$.

Algorithm 4.3. *Method due to Marsaglia and Tsang (2000) [63] for generating Gamma variates with $\alpha \geq 1$:*

- Setup $d = \alpha - \frac{1}{3}$, $c = 1/\sqrt{9d}$
- Generate $v = (1 + cx)$ where $(x \sim N(0, 1))$, and repeat this if $v \leq 0$
- Set $V = v^3$ and if $\log(U) < 0.5x^2 + d - dV + d\log(V)$ then return dV
- Go back to step 2.

It is possible to improve on the above algorithm by means of finding a promising 'squeeze function' which avoids the slightly more computationally expensive logarithms. The direct rejection method chooses a uniform point $(x, Ue^{-0.5x^2})$ under the curve $e^{-0.5x^2}$ and keeps x if $Ue^{-0.5x^2} < e^{g(x)}$. Applying a 'squeeze' amounts to finding some function $s(x)$ such that

$$s(x) \leq e^{g(x)} \leq e^{-0.5x^2} \tag{4.22}$$

and then accept x if $Ue^{-0.5x^2} \leq s(x)$. One such promising squeeze function is given by

$$s(x) = (1 - 0.0331x^4)e^{-0.5x^2} \tag{4.23}$$

In this form $s(x)$ is not that easy to evaluate; however, the exponential parts cancel in the rejection method's comparison, and only the $(1 - 0.0331x^4)$ part is important. The three curves $s(x), e^{g(x)}, e^{-0.5x^2}$ are very close over regions for which a normal x appears most of the time. It is worth noting that $s(x)$ is negative for $|x| > r = 0.0331^{-1/4}$ and so the squeeze is wasted in that region. For normal x, however, it is not very common to observe a value in that region. The squeeze ratio (area of s/ area $e^{g(x)}$) decreases ever so slightly for increasing α for the choice of 0.0331 and is about 0.9638 ($\alpha = 1$), 0.9199 (for $\alpha = 10$) and 0.91748 (for $\alpha = 100$). The tighter squeeze for smaller α partly compensates for the lower efficiency there, and this causes the final algorithm to have nearly constant average time for all $\alpha \geq 1$. A more formal proof that $s(x)$ is a squeeze function is given in [63].

1. Setup $d = \alpha - \frac{1}{3}$, $c = 1/\sqrt{9d}$
2. Generate $v = (1 + cx)$ where $(x \sim N(0, 1))$. Repeat this step if $v \leq 0$ - this is rare as it requires that $x < \sqrt{9\alpha - 3}$

3. Generate Uniform U and set $V = v^3$.

4. If $U < 1 - 0.0331x^4$ then return dV

5. If $\log(U) < 0.5x^2 + d(1 - V + log(V))$ then return dV

6. Go to step 2.

Marsaglia and Tsang go on to say that for the case where a gamma variate is required with $\alpha < 1$, the following relation may be used: $\gamma_\alpha = \gamma_{1+\alpha} U^{1/\alpha}$, where $U \sim$ Uniform[0,1].

The literature on gamma random variate generation is reasonably extensive [25, 36, 40] and a few more algorithms have been included:

Simulation algorithms with $\alpha \geq 1$

Algorithm 4.4. ***Best's generator*** *[13] of gamma variables when $a \geq 1$*

- Set $b = \alpha - 1$, $c = 3\alpha - 4$
- REPEAT
  - Generate i.i.d uniform [0,1] random variables U, V
  - Set $W = U(1 - U)$, $Y = \sqrt{\frac{c}{W}}\left(U - \frac{1}{2}\right)$, $X = b + Y$
  - If $X < 0$ go to REPEAT
  - Set $Z = 64W^3V^3 = (4WV)^3$
- UNTIL $\log(Z) \leq 2(b \log\left(\frac{X}{b}\right) - Y)$
- RETURN X

The following algorithm is a ratio of uniforms method due to Cheng and Feast [24]. The algorithm is slightly modified from its original form and appears as algorithm GKM1 in Glasserman.

Algorithm 4.5. *Method for sampling from gamma distribution for $(\alpha > 1)$.*

- Set $\bar{\alpha} = \alpha - 1$, $b = (\alpha - \frac{1}{6\alpha})/\bar{\alpha}$
- $m = \frac{2}{\bar{\alpha}}$, $d = m + 2$
- REPEAT
  - generate $U_1, U_2 \sim$ Unif[0,1]
  - $V = b\frac{U_2}{U_1}$
  - if $mU_1 - d + V + \left(\frac{1}{V}\right) \leq 0$, THEN accept
  - elseif $m \log U_1 - \log V + V - 1 \leq 0$, THEN accept
- UNTIL ACCEPT
- RETURN $Z = \bar{\alpha}V$

Simulation algorithms with $\alpha \leq 1$

Algorithm 4.6. ***Jöhnk's generator*** *[52] of gamma variables when $\alpha \leq 1$:*

- REPEAT
  - Generate i.i.d uniform [0,1] random variables U, V
  - Set $X = U^{1/\alpha}$, $Y = V^{1/(1-\alpha)}$
- UNTIL $X + Y \leq 1$
- Generate an exponential random variable, E
- RETURN $\frac{XE}{X+Y}$

The next algorithm is due to Ahrens and Dieter [4].

Algorithm 4.7. *Ahrens-Dieter method for generating gamma variables with $\alpha \leq 1$.*

- Set $e = \exp(1) = 2.71\ldots$ Set $b = \frac{a+e}{e}$.
- REPEAT
  - generate $U_1, U_2 \sim$ Unif[0,1], set $Y = bU_1$
  - IF $Y \leq 1$ then $Z = Y^{1/\alpha}$. Then, IF $U_2 < \exp(-Z) \Rightarrow$ ACCEPT
  - OTHERWISE set $Z = -\ln((b-Y)/\alpha)$. Then, IF $U_2 \leq Z^{\alpha-1} \Rightarrow$ ACCEPT
  - (Or, equivalently to above [36]: Generate $W_1 \sim$ Unif[0,1]. Set $W = W_1^{1/(\alpha-1)}$. Then, IF $W \geq Z \Rightarrow$ ACCEPT)
- UNTIL accept
- RETURN Z

The Ahrens-Dieter method can be modified as per the algorithm below [39] which has been included for the sake of completenes, even though it is not very different to the original algorithm.

Algorithm 4.8. *Variation of Ahrens-Dieter method*

- Set $t = 0.07 + 0.75\sqrt{1-\alpha}$ and $b = \alpha + \frac{e^{-t}\alpha}{t}$
- Generate U_1 and U_2 independently from Unif[0,1], and set $v = bU_1$
- IF $v \leq 1$ then
  - set $x = tv^{1/\alpha}$
  - if $U_2 \leq \frac{2-x}{2+x}$ then return x
  - otherwise, if $U_2 \leq e^{-x}$ then return x
- OTHERWISE
  - set $x = \log\left(\frac{t(b-v)}{\alpha}\right)$ and $y = \frac{x}{t}$
  - if $U_2(\alpha + y(1-\alpha)) \leq 1$ then return x
  - otherwise, if $U_2 \leq y^{a-1}$ return x

4.4.4 Sampling from the beta distribution

Proposition 4.1. If $X \sim \text{Gamma}(a, \beta)$ and $Y \sim \text{Gamma}(b, \beta)$, then the ratio

$$\frac{X}{X+Y} \sim \mathcal{B}(a, b) \tag{4.24}$$

Proof. First, recall that

$$B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{4.25}$$

Since X, Y are independent, their joint pdf is simply given by the product of their densities

$$f_{XY}(x, y) = \frac{\beta^{a+b} x^{a-1} y^{b-1} e^{-\beta(x+y)}}{\Gamma(a)\Gamma(b)} \tag{4.26}$$

Now, define $U = X/(X+Y)$ and $V = X+Y$. The inverse transformations are then $x = uv$ and $y = v(1-u)$. The Jacobian of this transformation is $|J| = v$. Now one simply needs to make use of a standard result from probability theory and transformations.

$$f_{UV}(u, v) = f(x, y)|J| = f(uv, v(1-u))v \tag{4.27}$$

$$= \frac{\beta^{a+b}(uv)^{a-1}(v(1-u))^{b-1} v e^{-\beta(v)}}{\Gamma(a)\Gamma(b)} \tag{4.28}$$

$$= \left(\frac{u^{a-1}(1-u)^{b-1}}{\Gamma(a)\Gamma(b)}\right)\left(v^{a+b-1} e^{-\beta(v)} \beta^{a+b}\right) \tag{4.29}$$

The first term is the density function (up to a normalising constant) for a Beta distribution with parameters (a, b), while the second term is the density function for a gamma distribution (also up to a normalising constant) with parameters $(a+b, \beta)$ (which makes sense since the gamma distribution is closed under convolution, thus we expect $X+Y = V$ to have gamma distribution). The fact that the joint function has been factorised into two marginal distributions indicates the independence of U and V. Thus the result is proved. □

This provides a relatively simple method for generating beta random variables for arbitrary parameters a and b. There is, however, a crucial dependence on the quality of the underlying gamma random variables. Inaccurate gamma random variates will lead to inaccurate (and hence unreliable) beta random variables. In most cases, for the purposes of this dissertation, when beta random variables are required, their parameters are identical ($a = b$) so they are sampled from a symmetric beta distribution. A highly accurate (double precision) means for inverting a symmetric beta distribution exists and is outlined below.

Inverting the symmetric beta distribution

L'Ecuyer and Simard [56] describe a method for inverting the cumulative distribution function of a symmetric Beta distribution in [56]. The method makes use of different series expansions for the CDF at different locations of $x \in [0, 1]$ aimed at optimising convergence. This approach combined with Newton's method to solve the equation $F(x) - u = 0$, yields a value of $x \in [0, 1]$ which follows a Beta distribution. The symmetric Beta distribution has a probability density function:

$$f(x) = \frac{[x(1-x)]^{\alpha-1}}{B(\alpha, \alpha)} \qquad (\forall x \in [0, 1]) \tag{4.30}$$

For purposes of sampling, first note that the symmetric Beta distribution satisfies the following relationship: $F(1-x) = 1 - F(x)$. Thus, it is only necessary to sample on $0 \leq x \leq 0.5$. For

this region, two different series representations of the Beta distribution are used, one for $x \simeq 0$ and another for $x \simeq 0.5$. The series representation also depends on the magnitude of α. The cumulative distribution function has integral representation:

$$F(x) = \frac{1}{B(\alpha, \alpha)} \int_0^x (t(1-t))^{\alpha-1} dt \tag{4.31}$$

When $0 < \alpha \leq 1$, $(1-t)^{\alpha-1}$ is replaced by its binomial series expansion, which is then integrated term by term. One obtains that, [56].

$$F(x) = \frac{1}{B(\alpha, \alpha)} \sum_{j=0}^{\infty} \frac{(1-\alpha)_j}{j!} \frac{x^{j+\alpha}}{j+\alpha} \tag{4.32}$$

where the $(b)_j$ term is Pochammer's symbol. This is defined as $\Gamma(b+j)/\Gamma(b)$ when b is not 0 or a negative integer. The above series has radius of convergence 1 and converges rapidly near $x = 0$, explaining its use over that region. When $x = 0.5$, each term is approximately half as big as the preceeding term, motivating the use of a different series in this region: Let $y = 0.5 - x$. Then

$$f(0.5 - x) = g(y) = \frac{(1/4 - y^2)^{\alpha-1}}{B(\alpha, \alpha)} \tag{4.33}$$

and, for $H(y) = 0.5 - F(0.5 - x)$ one obtains:

$$H(y) = \frac{y}{4^{\alpha-1} B(\alpha, \alpha)} \sum_{j=0}^{\infty} \frac{(1-\alpha)_j}{j!} \frac{(4y^2)^j}{2j+1} \tag{4.34}$$

This series has radius of convergence 0.5, converging rapidly for $y \simeq 0$ (i.e when x is close to 0.5). Thus, to compute $F(x)$ for a given value of x, use series (4.32) for $0 \leq x \leq x_m$ and series (4.34) when $x_m \leq x \leq 0.5$. These series are combined with Newton's method to obtain an accurate value for the Beta inverse function. It is shown by L'Ecuyer and Simard (2006) that one can use, for series (4.32), an initial estimate for Newton's method of

$$x_0 = (u\alpha B(\alpha, \alpha))^{1/\alpha} \tag{4.35}$$

Then, use the well known formula for subsequent iterates to obtain the following relation for iterating:

$$x_{n+1} = x_n - \left(\frac{F(x_n) - u}{f(x_n)} \right) \tag{4.36}$$

$$= x_n - \left(\sum_{j=0}^{\infty} \frac{(1-\alpha)_j x^{j+\alpha}}{j!(j+\alpha)} - uB(\alpha, \alpha) \right) \cdot [x(1-x)]^{1-\alpha} \tag{4.37}$$

Similarly, for series (4.34), the initial estimate is:

$$y_0 = \frac{w}{1 + 4(1-\alpha)w^2/3} \tag{4.38}$$

where,

$$w = 4^{\alpha-1} B(\alpha, \alpha)(0.5 - u) \tag{4.39}$$

The iteration formula then becomes ($v = 0.5 - u$):

$$y_{n+1} = y_n - \left(\frac{H(y_n) - v}{g(y_n)}\right) \tag{4.40}$$

$$= y_n \left(y_n \sum_{j=0}^{\infty} \frac{(1-\alpha)_j (4y^2)^j}{j!(j+\alpha)} - 4^{\alpha-1} B(\alpha,\alpha) v\right) \cdot [(1-4y^2)]^{1-\alpha} \tag{4.41}$$

This algorithm was coded in a Matlab function taking, as inputs, the 'symmetry' parameter α and a random number u in the interval [0,1]. The function returns the value, at that value of u, of the cumulative inverse beta distribution.

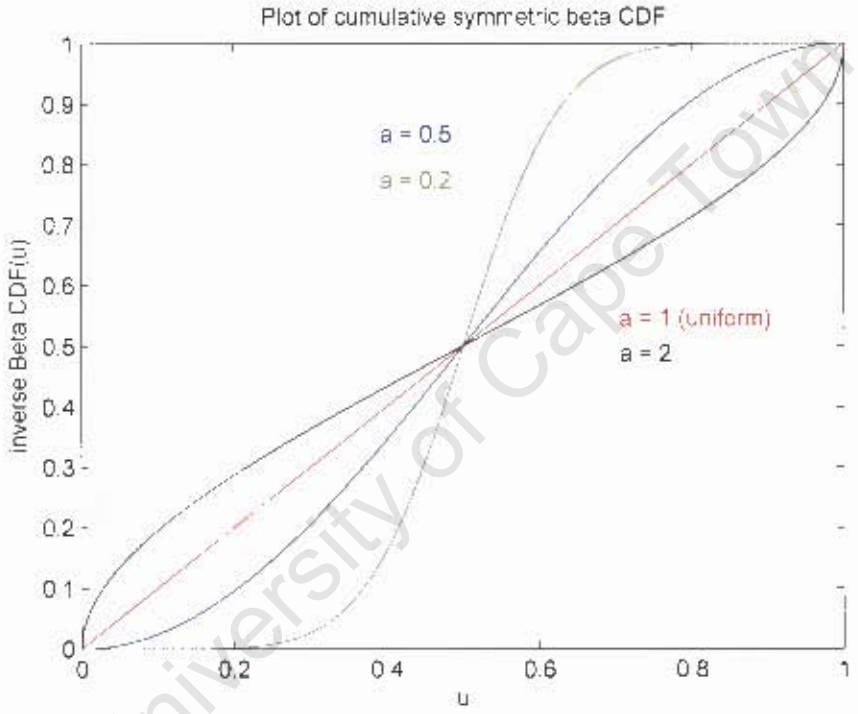

FIGURE 4.6: Plot of the inverse cumulative beta density for various values of the shape parameter using the betasyminv function [56].

4.4.5 Simulating from the inverse Gaussian distribution

The normal inverse Gaussian Lévy process, based on Brownian subordination, requires knowledge of simulation from the *inverse Gaussian* distribution. For this there is a remarkably simple and accurate algorithm due to Michael, Schucany and Haas [68]. If Y has the density given by equation (3.10) then:

$$\frac{(\gamma Y - \delta)^2}{Y} \sim \chi_1^2 \tag{4.42}$$

Therefore, one can sample Y by first sampling from the χ_1^2 distribution. This is a relatively straightforward task, since the χ_1^2 distribution is a special instance of the gamma distribution - it is $\mathcal{G}(1/2, 2)$. Alternatively, it is a well known fact that if $X \sim N(0,1)$ then $X^2 \sim \chi_1^2$, so good Gaussian or gamma generators are a prerequisite for inverse Gaussian random number

generation. Once a $\chi_1^2 = V$ random variable has been obtained, the resulting equation for Y has two roots (since it is quadratic in Y)

$$y_1 = \frac{\delta}{\gamma} + \frac{V}{2\gamma^2} - \frac{1}{2\delta\gamma}\sqrt{4\delta^3 V/\gamma + \delta^2 V^2/\gamma} \tag{4.43}$$

$$y_2 = \frac{\delta^2}{\gamma^2 y_1} \tag{4.44}$$

Michael, Schucany and Haas [68] show that the smaller root y_1 should be chosen with probability $\delta/(\delta+\gamma y_1)$ and the larger root y_2 with the complementary probability. Under the alternate (μ, λ) parameterisation, the above roots can be shown to correspond to:

$$y_1 = \mu + \frac{\mu^2 V}{2\lambda} - \frac{\mu}{2\lambda}\sqrt{4\lambda\mu V + \mu^2 V^2} \tag{4.45}$$

$$y_2 = \frac{\mu^2}{y_1} \tag{4.46}$$

The probability of accepting y_1 becomes $\mu/(\mu + y_1)$. The algorithm below can be found in Glasserman [40] and is consistent with the version in [25]. It was originally published in [68].

Algorithm 4.9. *Generating inverse Gaussian variables: [68]*

- Generate a normal random variable, N
- Set $V = N^2$
- Set $Y_1 = \mu + \frac{\mu^2 V}{2\lambda} - \frac{\mu}{2\lambda}\sqrt{4\mu\lambda V + \mu^2 V^2}$
- Generate uniform [0,1] random variable, U
- IF $U \leq \frac{\mu}{Y_1+\mu}$ RETURN Y_1, ELSE RETURN $\frac{\mu^2}{Y_1}$ $(= Y_2)$

Once an increment from the inverse Gaussian distribution has been obtained, it becomes possible to simulate an increment from the normal inverse Gaussian process since the process is defined to have normal increments subject to an inverse Gaussian time change.

4.4.6 Simulating α stable process trajectories

The standard simulation algorithm is due to Chambers, Mallows and Stuck [23]. The algorithm has particularly simple form for the symmetric case $\beta = 0$: if Y_1, Y_2 are independent random variables with Y_1 being standard exponential and Y_2 is uniformly distributed on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, then

$$X = \frac{\sin(\alpha Y_2)}{(\cos Y_2)^{1/\alpha}} \left(\frac{\cos((1-\alpha)Y_2)}{Y_1}\right)^{(1-\alpha)/\alpha} \tag{4.47}$$

has a $S_\alpha(1, 0, 0)$ distribution. Note, in the case that $\alpha = 2$ (Normal distribution), the form of X above reduces to:

$$X = 2\sqrt{Y_1}\sin(Y_2) \tag{4.48}$$

which is the Box-Muller algorithm for the generation of Gaussian random variates. It is possible to extend this result to general asymmetric stable distributions by means of the following result [7, 23]:
If Y_1 and Y_2 are both *i.i.d* $S_\alpha(1, 0, 0)$ variables, then

$$Y = \mu + \sigma\left(\frac{1+\beta}{2}\right)^{\frac{1}{\alpha}} Y_1 - \sigma\left(\frac{1-\beta}{2}\right)^{\frac{1}{\alpha}} Y_2 \tag{4.49}$$

has $S_\alpha(\sigma, \beta, \mu)$ distribution.

Chapter 5

Simulating Lévy process paths

Use of the techniques outlined in Chapter 4 are heavily relied upon for the simulation of Lévy process paths. A result, which is important for the simulation aspect of Lévy processes, is the Lévy -Ito decomposition theorem, because it states that *any* Lévy process can be written as a Brownian motion (with drift), an a.s finite sum of 'large' jumps and a possibly problematic sum of 'smaller' jumps which requires compensation in order that the sum remains finite. Simulating a Brownian motion is a straightforward matter, so provided that it is possible to calculate the two jump terms, it is possible to simulate the sample path of a Lévy process.

One property of Lévy processes which makes their simulation more interesting, is the presence of jumps in their price paths – these need to be accounted for somehow. In simulating jump-diffusion processes, the easiest way to incorporate the jumps is to firstly simulate the times when the jumps occur. Once one has the jump times and the jump size distribution, the cumulative size of the jump component for each jump time can be determined, and then this is simply added to the diffusion component. Simulating infinite-activity processes is a little bit more tricky, because the jumps happen far too often and are thus more difficult to keep track of. Fortunately, by simulating increments of the process directly from the distribution, it is possible to determine the value of the process on a fixed time grid. What happens in between time points is of less concern unless information is required in between two time points. In this case a finer grid should be used.

5.1 Compound Poisson Processes

Below, two algorithms are listed, which simulate sample paths of a compound Poisson process [25]. The second algorithm is an improvement on the first one from an implementation point of view. First the required number of jumps is determined by simulating from a Poisson distribution with a pre-specified intensity. Conditional on this number, N, the jump times are uniformly distributed, and are thus easy to simulate. The advantage here is that vector sizes can be assigned prior to simulation – when the program is being 'set-up'. It is worth mentioning though, that certain programming languages are more flexible than others when it comes to redimensionalising arrays, and so for instance in Matlab, this may not be much of an issue.

Algorithm 5.1. *Simulation of Compound Poisson Process (Algorithm 1)*

- *Initialise* $k = 0$
- *REPEAT WHILE* $\sum_{i=1}^{k} T_i < T$
- *set* $k := k + 1$

- *Simulate* $T_k \sim exp(\lambda)$
- *Simulate* Y_k *from distribution* $\mu = \nu/\lambda$

The trajectory is then given by: $X(t) = \gamma b + \sum_{i=1}^{N(t)} Y_i$, $N(t) = \sup\{k : \sum_{i=1}^{k} T_i \leq T\}$

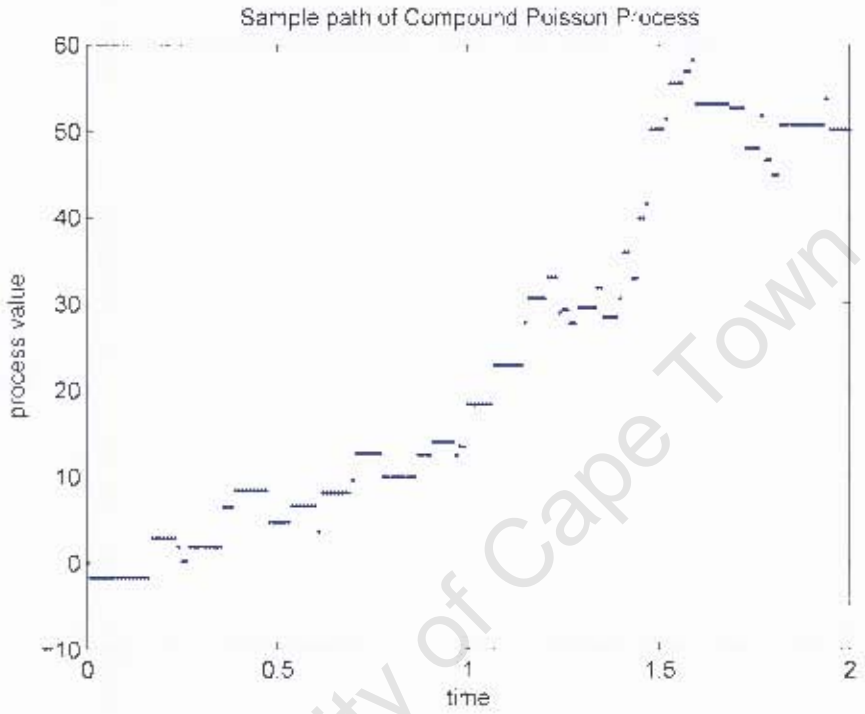

FIGURE 5.1: Sample path of compound Poisson process (no Gaussian component), $\lambda = 20$, jump distribution $\sim N(3, 1.5^2)$

Algorithm 5.2. *Simulation of Compound Poisson Process (Algorithm 2)*

- *Simulate random variable* N *from Poisson distribution with parameter* λT. *This is then the total number of jumps on interval* $[0, T]$.
- *Simulate* N *independent random variables* U_i, *uniformly distributed on* $[0, T]$. *These are the jump times.*
- *Simulate jump sizes:* N *independent random variables* Y_i *with law* $\nu(dx)/\lambda$

The trajectory is then given by:

$$X(t) = bt + \sum_{i=1}^{N} I_{\{U_i \leq t\}} Y_i \tag{5.1}$$

Figure 5.1 illustrates a typical sample path of a CPP. A Poisson process with an intensity parameter $\lambda = 20$ was used, and the jump size distribution was normal with parameters $\mu = 3, \sigma = 1.5$. To illustrate a typical sample path of a jump-diffusion process (Figure 5.2), exactly the same process was used, with a Gaussian component included. This particular case is simply the Merton model [67].

5.2 Simulation of Jump-Diffusions on a fixed time grid

This algorithm formalises the discussion presented in the introduction section of this chapter and is taken from [25]:

Algorithm 5.3. *Simulation of* $(X_1, \ldots, X_n)$ *for* n *fixed times,* $t_1, \ldots, t_n$

- Simulate n independent centered Gaussian random variables G_i with variances $\mathrm{Var}(G_i) = (t_i - t_{i-1})\sigma^2$, where $t_0 = 0$.
- Simulate the Compound Poisson part by Algorithm 5.2

The discretized trajectory is given by:

$$X(t_i) = bt_i + \sum_{k=1}^{i} G_k + \sum_{j=1}^{N} I_{\{U_j \leq t_i\}} Y_j \tag{5.2}$$

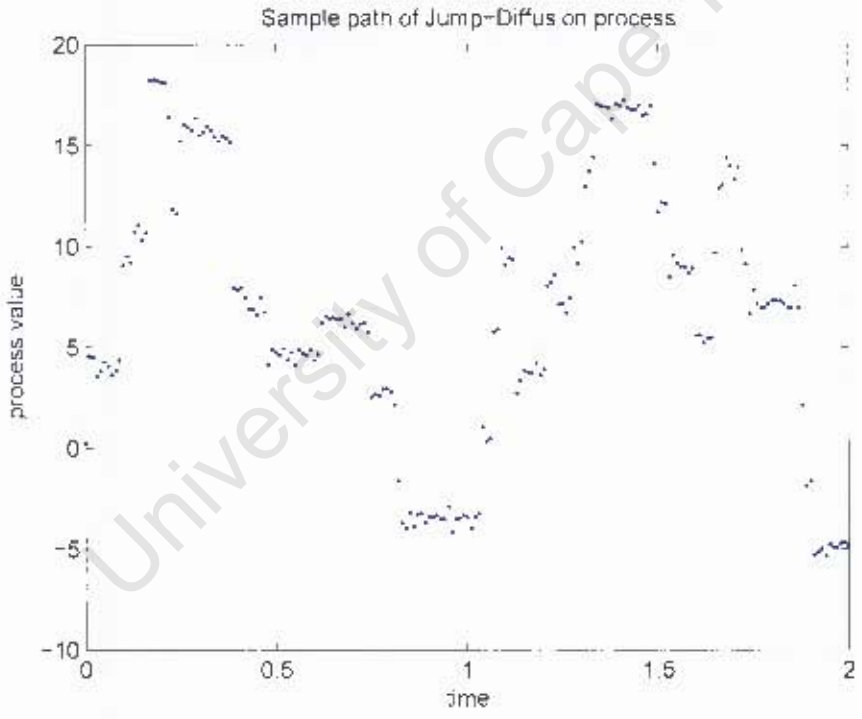

FIGURE 5.2: Sample path of jump-diffusion process (CPP + standard Gaussian component) $\lambda = 20$, jump distribution $\sim N(3, 1.5^2)$

5.3 Exact simulation of increments and Brownian subordination

The algorithms presented below aim to simulate actual increments from the Lévy process itself. In moving away from jump-diffusions this approach becomes useful because it does not depend on knowing exactly when the process jumps. A familiar result is stated below which is used extensively for modelling in the Black-Scholes framework:

5.4.1 Sequential sampling

Sequential sampling is a sampling methodology whereby the calculation of the next value in the sequence is based on the previous value obtained. The process is assumed to start at zero whereafter the next increment is simulated. The algorithm thus constructs sample paths in a forward looking manner; it does not assume that anything about the process is known beyond the current time t_i. The two processes G_t (the gamma process) and X_t (the Brownian motion) are assumed to be independent. The value at time t_i is used to calculate the value at time t_{i+1} as outlined below:

Algorithm 5.7 (VGSS). *Simulation of a Variance-Gamma process on a fixed time grid: [25] Simulation of $(X_1, \ldots, X_n)$ for fixed times, $t_1, \ldots, t_n$, a discretized trajectory of the Variance-Gamma process with parameters σ, θ, κ.*

- Simulate, using Algorithms 4.6 and 4.4, n independent Gamma variables: $\Delta S_1, \ldots, \Delta S_n$ with mean parameters $\frac{t_1}{\kappa}, \frac{t_2 - t_1}{\kappa}, \ldots, \frac{t_n - t_{n-1}}{\kappa}$ and variance parameter $= 1$. Set $\Delta S_i = \kappa \Delta S_i \quad (\forall i)$.
- Simulate n i.i.d N(0,1) random variables $N_1, \ldots, N_n$ and set $\Delta X_i = \sigma N_i \sqrt{\Delta S_i} + \theta \Delta S_i$

then,

$$X(t_i) = \sum_{k=1}^{i} \Delta X_k$$

5.4.2 Brownian-Gamma Bridge Sampling

In contrast to sequential sampling, bridge sampling concentrates more on the so called 'macro-effects' (which, in a nutshell, is sampling over longer time intervals). The method assumes that the starting value of the process is zero, and uses the distributional properties of the process at time T to simulate a value for the process at time T. This is possible because the closed form conditional distributions are available. Subsequent generation of time points is done in a specific order, as follows: T, $T/2$, $T/4$, $3T/4$, $T/8$, $3T/8$, $5T/8$, $7T/8$ and so on. For bridge sampling, the following result is crucial because it gives the distribution of the process conditional on the end points:

Proposition 5.1. Let G_t be a gamma process on the interval [0,T], where $G_t \sim \mathcal{G}\left(\frac{t\mu^2}{\nu}, \frac{\mu}{\nu}\right)$. Suppose there are times t_1 and t_2 such that $0 < t_1 < \tau < t_2 < T$ for which the process is known (i.e $G_{t_1} = \gamma_1$ and $G_{t_2} = \gamma_2$). Then the distribution of G_τ conditional on $G_{t_1} = \gamma_1$ and $G_{t_2} = \gamma_2$ is equal in distribution to:

$$G_\tau \sim G_{t_1} + (G_{t_2} - G_{t_1})Y \tag{5.5}$$

Where $Y \sim \mathcal{B}\left(\frac{(\tau - t_1)\mu^2}{\nu}, \frac{(t_2 - \tau)\mu^2}{\nu}\right)$.

Proof. Since $(G_t)_{t \geq 0}$ has independent increments,

$$G_\tau - G_{t_1} \sim \mathcal{G}\left(\frac{(\tau - t_1)\mu^2}{\nu}, \frac{\mu}{\nu}\right)$$

$$G_{t_2} - G_\tau \sim \mathcal{G}\left(\frac{(t_2 - \tau)\mu^2}{\nu}, \frac{\mu}{\nu}\right)$$

are independent gamma variates. Thus the ratio (Proposition 4.1):

$$\frac{G_\tau - G_{t_1}}{G_\tau - G_{t_1} + G_{t_2} - G_\tau} = \frac{G_\tau - G_{t_1}}{G_{t_2} - G_{t_1}} = Y \tag{5.6}$$

follows a Beta distribution with parameters $(\tau - t_1)\mu^2/\nu$ and $(t_2 - \tau)\mu^2/\nu$. As our interest lies with the conditional distribution, G_τ, G_{t_1} and G_{t_2} are but mere constants, and thus rewriting the above for G_τ gives the result, i.e:

$$G_\tau = G_{t_1} + (G_{t_2} - G_{t_1})Y$$

□

In particular, if τ is the midpoint between t_1 and t_2, then the beta distribution in question is symmetric and efficient methods exist for inverting the symmetric beta distribution [56].

For this sampling technique, again, G_t and X_t are assumed to be independent. This implies that conditional on any collection of increments of the gamma process, the collection of increments of the Brownian motion will be independent normals. All this means, quite simply, is that increments from G can be sampled followed by increments of $B(G(t))$ by means of Brownian bridge sampling (conditional on the corresponding G increments). Either, one can sample all the increments of G followed by all the increments of B, or one can sample them in alternance. The latter method is used below:

Algorithm 5.8 (BGBS). *Simulation of a variance-gamma process:*$X(t) = B(G(t;\mu,\nu),\theta,\sigma)$ *for a* 2^k *equal-length partition of [0,T]*

- Set $G(0) = 0$ and $X(0) = 0$
- Generate $G(T) \sim \mathcal{G}(\frac{\mu^2 T}{\nu}, \frac{\mu}{\nu})$ and $X(T) \sim \mathcal{N}(\theta G(T), \sigma^2 G(T))$
- For $l = 1$ to k
  - For $m = 1$ to 2^{l-1}
    - $i = 2m - 1$
    - Generate $Y \sim \mathcal{B}(\frac{\mu^2 T}{\nu 2^l}, \frac{\mu^2 T}{\nu 2^l})$
    - $G\left(i\frac{T}{2^l}\right) = G\left((i-1)\frac{T}{2^l}\right) + \left[G\left((i+1)\frac{T}{2^l}\right) - G\left((i-1)\frac{T}{2^l}\right)\right] Y$
    - $b = G\left((i+1)\frac{T}{2^l}\right) - G\left(i\frac{T}{2^l}\right)$
    - Generate $Z \sim N(0, b\sigma^2 Y)$
    - $X\left(i\frac{T}{2^l}\right) = Y \cdot X\left((i+1)\frac{T}{2^l}\right) + (1-Y)\cdot X\left((i-1)\frac{T}{2^l}\right) + Z$
  - Next m
- Next l

5.4.3 Double-Gamma Bridge Sampling

This particular algorithm exploits the alternate representation of a variance-gamma process as the difference of two increasing (gamma) processes [19]. One writes the variance-gamma process as

$$X(t;\theta,\sigma,\nu) = \gamma_p(t;\mu_p,\nu_p) - \gamma_n(t;\mu_n,\nu_n) \tag{5.7}$$

where (courtesy of relations in (3.51)):

$$\mu_p = \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} + \frac{\theta}{2} \qquad \mu_n = \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} - \frac{\theta}{2} \tag{5.8}$$

$$\nu_p = \mu_p^2\nu \qquad \nu_n = \mu_n^2\nu \tag{5.9}$$

Algorithm 5.9 (DGBS). *Simulation of a variance-gamma process $X(t) = B(G(t;\mu,\nu),\theta,\sigma)$ for a 2^k equal-length partition of $[0,T]$*

- Set $\gamma_p(0) = 0$ and $\gamma_n(0) = 0$
- Generate $\gamma_p(T) \sim \mathcal{G}(\frac{\mu_p^2 T}{\nu_p}, \frac{\mu_p}{\nu_p})$ and $\gamma_n(T) \sim \mathcal{G}(\frac{\mu_n^2 T}{\nu_n}, \frac{\mu_n}{\nu_n})$
- For $l = 1$ to k
  - For $m = 1$ to 2^{l-1}
    - $i = 2m - 1$
    - Generate $Y_p \sim \mathcal{B}(\frac{\mu_p^2 T}{\nu_p 2^l}, \frac{\mu_p^2 T}{\nu_p 2^l})$
    - $\gamma_p\left(i\frac{T}{2^l}\right) = \gamma_p\left((i-1)\frac{T}{2^l}\right) + \left[\gamma_p\left((i+1)\frac{T}{2^l}\right) - \gamma_p\left((i-1)\frac{T}{2^l}\right)\right] Y_p$
    - Generate $Y_n \sim \mathcal{B}(\frac{\mu_n^2 T}{\nu_n 2^l}, \frac{\mu_n^2 T}{\nu_n 2^l})$
    - $\gamma_n\left(i\frac{T}{2^l}\right) = \gamma_n\left((i-1)\frac{T}{2^l}\right) + \left[\gamma_n\left((i+1)\frac{T}{2^l}\right) - \gamma_n\left((i-1)\frac{T}{2^l}\right)\right] Y_n$
    - $X\left(i\frac{T}{2^l}\right) = \gamma_p\left(i\frac{T}{2^l}\right) - \gamma_n\left(i\frac{T}{2^l}\right)$
  - Next m
- Next l

5.5 Simulation algorithm for NIG sample paths

The process described below is almost identical to Algorithm 5.4.1, except the subordinator increments are sampled from the inverse Gaussian distribution and not the gamma distribution. The algorithm is as follows:

Algorithm 5.10. *Simulation of NIG Process (Normal Inverse Gaussian) on a fixed time grid:[25] Simulation of $(X_1,\ldots,X_n)$ for fixed times, $t_1,\ldots,t_n$, a discretized trajectory of a NIG process with parameters σ,θ,κ*

- Simulate, using 4.9 n independent Inverse Gaussian variables, $\Delta S_1,\ldots,\Delta S_n$ with parameters $\lambda_i = \frac{(t_i - t_{i-1})^2}{\kappa}$ and $\mu_i = t_i - t_{i-1}$, where t_0 is taken to be 0
- Simulate n i.i.d N(0,1) random variables $N_1,\ldots,N_n$ and set $\Delta X_i = \sigma N_i\sqrt{\Delta S_i} + \theta\Delta S_i(\quad \forall i)$

The (discretized) trajectory is given by (5.3).

5.6 Approximations of Lévy processes as Compound Poisson Processes

One useful consequence of the Lévy -Ito decomposition theorem (Theorem 2.2), is that any Lévy process can be written as the sum of a continuous Gaussian component, a compound Poisson process – for jumps which are big enough – and a compensated CPP for the jumps which are

small . In particular, if X_t is a Lévy process with characteristic triplet $(0, \nu, \gamma)$ then it has representation:

$$X_t = \gamma_t + \sum_{s \leq t} \Delta X_s \mathbb{I}_{\{|\Delta X_s| \geq 1\}} + N_t^{\varepsilon} \tag{5.10}$$

where

$$N_t^{\varepsilon} = \sum_{s \leq t} \Delta X_s \mathbb{I}_{\{\varepsilon \leq |\Delta X_s| < 1\}} - t \int_{\varepsilon \leq |x| \leq 1} x\nu(dx) \tag{5.11}$$

It turns out that the residual process – defined by $R_t^{\varepsilon} = \lim_{\delta \downarrow 0} N_t^{\delta} - N_t^{\varepsilon}$ is a Lévy process [25] – with characteristic triplet $(0, \mathbb{I}_{|x| \leq \varepsilon}\nu(dx), 0)$ and $\mathbb{E}[R_t^{\varepsilon}] = 0$. In the case where this approximation is applied to finite variation Lévy processes, small jumps are not problematic enough and do not need to be compensated for; they can simply be replaced by their expectation. These processes can now be simulated by means of a compound Poisson simulation algorithm. The quality of this approximation depends on how $\sigma^2(\varepsilon) \to 0$ as $\varepsilon \to 0$. This is because the error process is an infinite activity process with bounded jumps, and therefore has finite variance [25]. Define $\sigma^2(\varepsilon)$ as follows:

$$\sigma^2(\varepsilon) = \mathrm{Var}[R_t^{\varepsilon}] = t \int_{|x| \leq \varepsilon} x^2 \nu(x) dx \tag{5.12}$$

The compound Poisson approximation is thus a good approximation provided that there are not too many small jumps, or equivalently, if the jumps are reasonably well behaved in a neighbourhood of zero. Said another way, it means that the growth of the Lévy measure should not be too fast in this region.

An alternative approach is to make use of a (possibly infinite) number of independent Poisson processes added together, to approximate a Lévy process as laid out in Schoutens (2003) [76]. The methodology is as follows: Suppose (X_t) is a $(\sigma^2, \nu(dx), \gamma)$ Lévy process. The first step is deciding how many Poisson processes to use in the approximation. The Lévy measure is then discretised (divided into regions or 'bins') governed by this number. For instance, deciding to use $d = 2k$ Poisson processes means that there can be an equal number of positive and negative approximating processes, each representing different jump size intensities corresponding to a specific interval. The discretisation is achieved through a partition, for some $\varepsilon \in \mathbb{R}$, of $\mathbb{R} \backslash [-\varepsilon, \varepsilon]$. This shall be referred to as $\{a_0, a_1, \ldots, a_k = -\varepsilon, \epsilon = a_{k+1}, \ldots, a_{d+1}\}$. It is not a necessary requirement that the number of positive and negative intervals are the same, but this is often simplest to implement. Due to the nature of most Lévy measures, which have a singularity at the origin (if the process is infinite activity) it is often best to avoid using 'quick and dirty' techniques to select the partition values. A few approaches are suggested and discussed below:

- **Equally spaced intervals:** Here, the lengths of the partition intervals, $|a_i - a_{i-1}|$, are kept constant. This is easy to set up, but the nature of the Lévy measure typically results in an 'explosion' in the region of 0, which is problematic.

- **Equally weighted intervals:** Another approach is to keep the Lévy measure of each partition interval, $\nu([a_{i-1}, a_i))$, constant. The partition values are calculated so that the up jump and down jump intensities is the same throughout. One problem with this method is that the outer most partition values can become very large. Also, the calculation of partition values in this case is computationally intensive because one has to integrate the Lévy measure, while varying the boundaries until the integral value converges to the chosen constant value. Newton's method may be required to achieve convergence quickly.

- **Intervals with inverse linear boundaries:** Here, for $1 \leq i \leq k+1, \alpha > 0$, set:

$$a_{i-1} = -\alpha i^{-1} \tag{5.13a}$$

$$a_{2k+2-i} = \alpha i^{-1} \tag{5.13b}$$

  An advantage of using this interval construction is that there is no explosion to infinity in the region of zero, and there is even a tapering off the intensities. This was the chosen method for the purposes of this dissertation.

With the choice of interval pinned down, on each interval $[a_{i-1}, a_i)$, define an independent Poisson process (to those associated with other intervals), $N^{(i)} = \{N_t^{(i)}, t \geq 0\}$. The parameters of each Poisson process used in the approximation of $(X_t)_{t\geq 0}$, denoted by λ_i and c_i respectively, represent the intensity and jump size of the corresponding Poisson process used in the approximating process, $(X_t^d)_{t\geq 0}$. Then [76],

$$X_t^d = \gamma_t + \sigma W_t + \sum_{i=1}^{d} c_i (N_t^{(i)} - \lambda_i t \mathbb{I}_{\{|c_i|<1\}}) \tag{5.14}$$

The parameter λ_i is chosen to be equal to the Lévy measure on the corresponding interval, while c_i is chosen such that the variance of the approximating process matches the variance of the Lévy process

$$\lambda_i = \begin{cases} \nu([a_{i-1}, a_i)) & \text{for } 1 \leq i \leq k \\ \nu([a_i, a_{i+1})) & \text{for } k+1 \leq i \leq d \end{cases} \tag{5.15}$$

$$c_i^2 \lambda_i = \begin{cases} \int_{a_{i-1}}^{a_i^-} x^2 \nu(dx) & \text{for } 1 \leq i \leq k \\ \int_{a_i}^{a_{i-1}^-} x^2 \nu(dx) & \text{for } k+1 \leq i \leq d \end{cases} \tag{5.16}$$

The integrals of the Lévy measure were calculated using an integration technique known as *Romberg integration*. This quadrature technique is similar to the trapezium rule, but it is an inductive procedure that is straightforward to code up efficiently. Sample Matlab code was available from the Mathworks website [54].

Finally, the small jumps need to be considered and dealt with. When there sufficiently many small jumps, to the extent that they are problematic, then they need to be replaced with a Brownian motion. This correction ends up being more efficient than the compound Poisson approximation [25]. A theorem by Asmussen and Rosinski (2002) provides a condition which validates the use of a Brownian motion approximation. The theorem states that $\sigma(\varepsilon)^{-1} R^\varepsilon \to W$ *in distribution as* $\varepsilon \to 0$ *iff for all* $k > 0$ *we have that as* $\varepsilon \to 0$,

$$\frac{\sigma(k\sigma(\varepsilon) \wedge \varepsilon)}{\sigma(\varepsilon)} \to 1 \tag{5.17}$$

This condition is implied by $\sigma(\varepsilon)/\varepsilon \to \infty$ as $\varepsilon \to 0$ – often easier to check [25, 76]. The following examples give examples of processes which may make use of a Brownian motion approximation, and can all be found in Schoutens (2003) [76]:

Example 5.1 (NIG).

$$\sigma(\varepsilon) \sim \sqrt{\frac{2\alpha\delta}{\pi}} \varepsilon^{\frac{1}{2}} \tag{5.18}$$

Condition is satisfied.

Example 5.2 (Meixner).

$$\sigma(\varepsilon) \sim \text{NIG} \tag{5.19}$$

Condition is satisfied.

Example 5.3 (CGMY).

$$\frac{\sigma(\varepsilon)}{\varepsilon} \to \infty \iff Y > 0 \tag{5.20}$$

Example 5.4 (Gamma).

$$\frac{\sigma(\varepsilon)}{\varepsilon} \to \sqrt{\frac{\alpha}{2}}, \quad \varepsilon \to 0 \tag{5.21}$$

Example 5.5 (VG). Since this process is the difference of two gamma processes, and these processes do not permit small jumps to be approximated by a Brownian motion, this approximation is not valid for a variance-gamma process either.

5.7 Modelling in higher dimensions – dependence and copulas

An advantage of and reason for the ubiquity of multivariate Gaussian models in the literature is because extension into a multivariate setting is simple. An n dimensional multivariate normal distribution can be specified through two things: n normal marginal distributions (this is important) and $\frac{n}{2}(n-1)$ correlations – one between each pair of margins. This is enough to completely classify the dependence structure between all the variables. One problem with this approach is that it has limited flexibility. It does allow you to directly specify the correlation between any two of the variables, but the overall structure of this dependence is limiting.

It may be argued, in the context of Lévy processes, that extension of the subordinator based models, should be simple: simply time change a multivariate Brownian motion with a univariate subordinator. In doing this however, one is limited to a narrow range of dependence structures which excludes total independence and restrict all marginal components to be of the same type. Suppose that a two-dimensional Brownian motion is used to model two correlated stock price processes:

$$\begin{aligned} S_t^1 &= \exp(X_t^1), & X_t^1 &= B^1(Z_t) + \mu_1 Z_t \\ S_t^2 &= \exp(X_t^1), & X_t^2 &= B^2(Z_t) + \mu_2 Z_t \end{aligned}$$

The correlation of returns is given [25] as

$$\rho(X_t^1, X_t^2) = \frac{\sigma_1\sigma_2\varrho\mathbb{E}[Z_t] + \mu_1\mu_2\text{Var}[Z_t]}{(\sigma_1^2\mathbb{E}[Z_t] + \mu_1^2\text{Var}[Z_t])^{1/2}(\sigma_2^2\mathbb{E}[Z_t] + \mu_2^2\text{Var}[Z_t])^{1/2}} \tag{5.22}$$

Cont & Tankov (2004) argue that even in a symmetric case – when the two Brownian motions are independent – that since the processes are being subordinated by the same subordinator, there will be a correlation in the absolute values of returns. This is because large jumps in the stock prices will tend to arrive together. When $\mu_1 = \mu_2 = 0$ and $\varrho = 0$ the covariance of the squares of returns is [25]

$$\text{Cov}[(X_t^1)^2, (X_t^2)^2] = \sigma_1\sigma_2\text{Cov}[W_1^2(Z_t), W_2^2(Z_t)] = \sigma_1\sigma_2\text{Var}[Z_t]$$

and thus squares of returns are correlated if Z_t is not deterministic.

One way to overcome these limitations is to make use of a *copula*. In simple terms, a *copula* is a

function satisfying two conditions: It must be a multivariate distribution function, and it must have margins which are all uniform. If one can find such a function, then it becomes possible to form a multivariate distribution with a specific dependence structure. This dependence structure is governed by the choice of copula function. The theorem which states how the copula and marginals define a multivariate distribution is Sklar's theorem. Again, no formal definition will be given here, but the theorem states that for a joint (n-dimensional) distribution function, with n marginals $F_1, \ldots, F_n$, there exists an n-dimensional copula C such that for all $x \in \bar{\mathbb{R}}^n$, one has

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \tag{5.23}$$

Thus, it is essentially an existence theorem for copulas. It goes on to say that, if the marginal functions are all continuous then C is unique. Conversely, given a multivariate distribution, F, with marginals $F_1, \ldots, F_n$, setting $C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n))$ shows that every multivariate distribution admits the representation in (5.23) [25, 40].

Suppose, in a naïeve setting, there was a way to impose a multivariate Gaussian type of structure to non Gaussian marginal distributions. Investigating this would be worthwhile as it would help to create modelling flexibility. For example, suppose one wanted to simulate dependent random variables with arbitrary marginal distributions, from variables following a standard bivariate normal distribution. The following could be done:

1. Simulate two vectors (Z_1 and Z_2) of length n from a standard bivariate normal distribution with correlation coefficient ρ.
2. Apply the normal CDF function to each element of these 2 vectors, so as to obtain two vectors of uniformly distributed random variables. $U_1 = \Phi(Z_1)$ and $U_2 = \Phi(Z_2)$, where Φ denotes the normal CDF.
3. Apply inverse CDF operations to each of the 2 vectors to transform them to the desired marginal distributions. $X_1 = G_1^{-1}(U_1)$ and $X_2 = G_2^{-1}(U_2)$

It must be pointed out that the correlation coefficient between X_1 and X_2 will no longer be ρ since Pearsons correlation (the standard) is not invariant under even a strictly increasing transformation. This is why it is often better to use an alternate measure of correlation such as Spearmans rank correlation or Kendal's tau coefficient since these are constructed in a way which makes them invariant under strictly increasing transformations. One could thus, for example, pick G_1^{-1} to be the inverse function of a gamma distribution with parameters 3 and 2 and G_2^{-1} to be the inverse function of a students-t distribution with parameter 5. This is the essence of modelling with copulas.

There are many different copulas which can be read about and used, but problems arise when one tries to incorporate these for use with Lévy processes. This is illustrated by the following example: Consider a pure jump Lévy process given by (X_t) and suppose that from the jumps of (X_t) another Lévy process is formed: $(Y_t) = \sum_{s \leq t} (\Delta X_s)^3$. It seems plausible to claim that (X_t) and (Y_t) are fully dependent, since the trajectory of one of them can be constructed from the trajectory of the other. However, the copula joining (X_t) and (Y_t) is not one of complete dependence since (Y_t) is not a deterministic function of (X_t). A definition of when two Lévy processes are independent is required:

Definition 5.1 (Independent Lévy processes). *Let (X_t, Y_t) be a Lévy process with Lévy measure ν and no Gaussian component. The components of this Lévy process are independent iff the support of ν is contained in the set $\{(x, y) : xy = 0\}$. This happens iff the two processes never jump together, in which case:*

$$\nu(A) = \nu_X(A_X) + \nu_Y(A_Y) \tag{5.24}$$

where, $A_X = \{x : (x, 0) \in A\}$, $A_Y = \{y : (0, y) \in A\}$, *and* ν_X *and* ν_Y *are Lévy measures of* (X_t) *and* (Y_t) *respectively.*

If the two Lévy processes never jump together and this has implications for their Lévy measures, then these Lévy measures must play a defining role in constructing an equivalent notion to that of a copula for Lévy processes. These are known as *Lévy copulas.* This dissertation will not consider the issue of modelling dependence of Lévy processes. The Lévy copula will be defined, but the reader will be referred to sources such as Cont & Tankov [25] and Tankov [79]. Further research extending the ideas implemented in this dissertation could be undertaken in this area.

Definition 5.2 (Lévy copula)**.** A function $F : \bar{\mathbb{R}}^d \mapsto \mathbb{R}$ is called a Lévy copula if [25]

1. $F(u_1, \ldots, u_d) \neq \infty$ for $(u_1, \ldots, u_d) \neq (\infty, \ldots, \infty)$
2. $F(u_1, \ldots, u_d) = 0$ if $u_i = 0$ for at least one $i \in \{1, \ldots, d\}$
3. F is d-increasing
4. $F^{\{i\}}(u) = u$ for any $i \in \{1, \ldots, d\}, u \in \mathbb{R}$

where for any $I \in \{1, \ldots, d\}$, $F^{\{i\}}(u)$ is the $I-$*margin* of F. This is a function from $\bar{\mathbb{R}}^I \mapsto \mathbb{R}$ defined by

$$F^I((u_i)_{i \in I}) = \lim_{c \to \infty} \sum_{(u_j)_{j \in I^c} \in \{-c, \infty\}} F(u_1, \ldots, u_d) \prod_{c \in I^c} \text{sgn} u_j \tag{5.25}$$

and $I^c = \{1, \ldots, d\}/I$.

Chapter 6

Numerical Results

Chapter 4 mapped out the methodology for simulating random variables from a wide variety of different distributions, with Chapter 5 discussing various methods of simulating Lévy processes. The focus of this chapter is on practical aspects of the implementation of these algorithms. Practically, speed and accuracy of algorithms is an important consideration in the development of any simulation based model. Desirable models include those which are not only as accurate

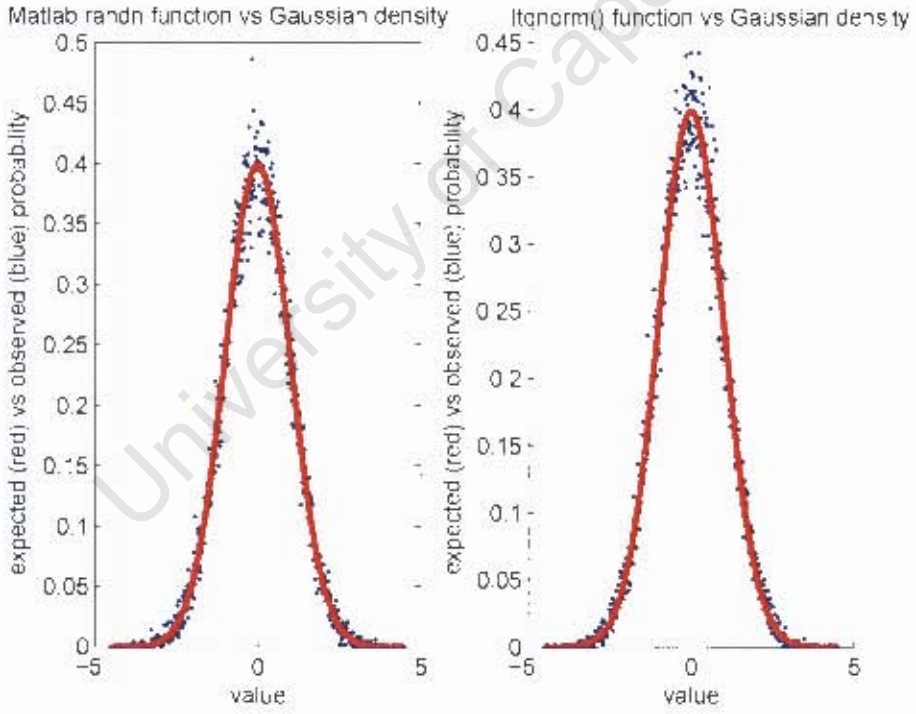

FIGURE 6.1: Comparing the normal random number generators of Matlab (built in - left) and that of a function given by Peter Acklam (right)

as possible, but which are also able to produce results in a reasonable turnaround time. Waiting days for a numerical result in a fast moving and dynamic market environment is less than ideal, and thus ways of speeding up the modelling process are certainly worth looking into.

This dissertation will not be concerned with numerical and time-based optimisation of simulation code – although it could certainly be looked into. The optimisation aspect of coding is very

platform specific with each programming language having its own strengths and weaknesses. The aim here, rather, is to give a better, more generic sense and feel for whether the given algorithms are accurate and reliable for use in a modelling context. Also, in instances where a handful of algorithms are available it will aim to argue for the best method and motivate this choice. There, in many cases, may be refinements that would reduce the computational time significantly but these have not been considered. The simulations were run on an *Intel Pentium*

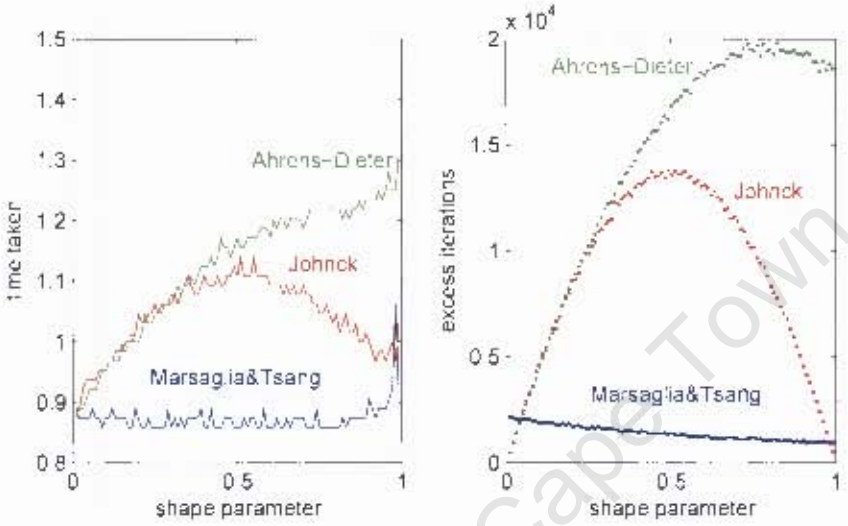

FIGURE 6.2: Figures showing the time taken (left) and number of excess iterations (above 50,000 - right) for the various sampling techniques when the shape parameter is less than 1

M processor, with a clock speed of 1.50 GHz and 512MB of RAM, in either a Matlab, C or an Excel environment. A discussion of the implementation of uniform and normal random number generation used in the simulations of Lévy processes follows.

The standard RAND function in Matlab used for simulations was the *Mersenne Twister* [65]. The computational time taken to generate 50000 random variables in Matlab was not even computationally significant (this registered as 0.000 seconds), and given the properties of this random number generation algorithm, it was used as a platform for all uniform pseudo-random number generation. Normal random variables were generated using Matlab's RANDN function. This is a built in function with no given references in the function's documentation. It is also able to generate 50000 variables in no time at all. On top of this it is accurate in that the first four central moments agree with the theoretical quantities and the histogram is a fairly good representation of the normal density function (Figure 6.1). It was as accurate as Peter Acklam's normal inverse function as well as the Polar-Marsaglia method (see Section 4.4.1). When tested formally using the Kolmogorov-Smirnov test for normality in Matlab, even at a 0.1% significance level, the null hypothesis that the data are normally distributed using either method was not rejected.

6.1 Simulating gamma variates

A simple study of the generation time taken to obtain 50000 gamma variates, using each of the particular (and relevant) algorithms for different shape parameters of the gamma distribution, was carried out. Since all algorithms were acceptance / rejection based, the number of loops required to obtain the 50000 variates was also recorded and included in the study. Figures 6.2

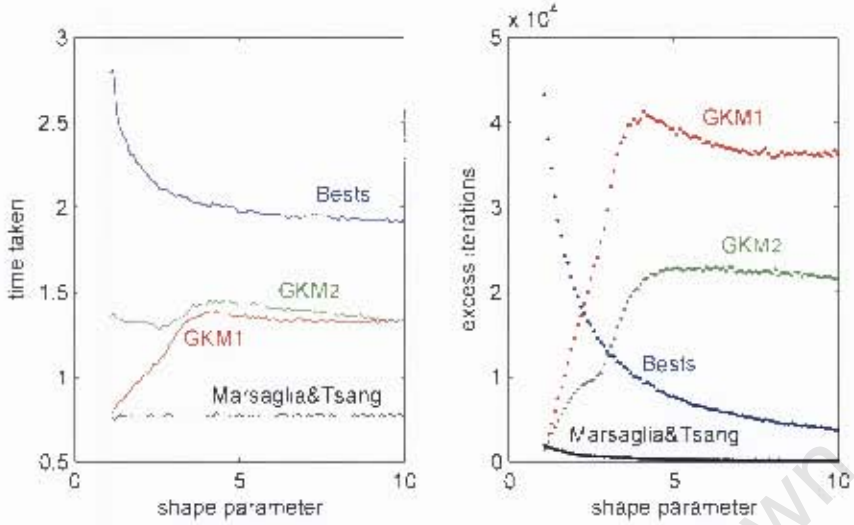

FIGURE 6.3: Figures showing the time taken (left) and number of excess iterations (right) for the various sampling techniques where the shape parameter is between 1 and 10

through 6.4 all contain two plots. The plot on the left is of the time taken for each particular algorithm, given a value of the shape parameter, while the graph on the right is of the number of excess iterations required for that particular shape parameter. In other words, it is a graph of the number of rejected random numbers.

A brief glance at Figures 6.2 - 6.4 indicates that in terms of all round performance the method

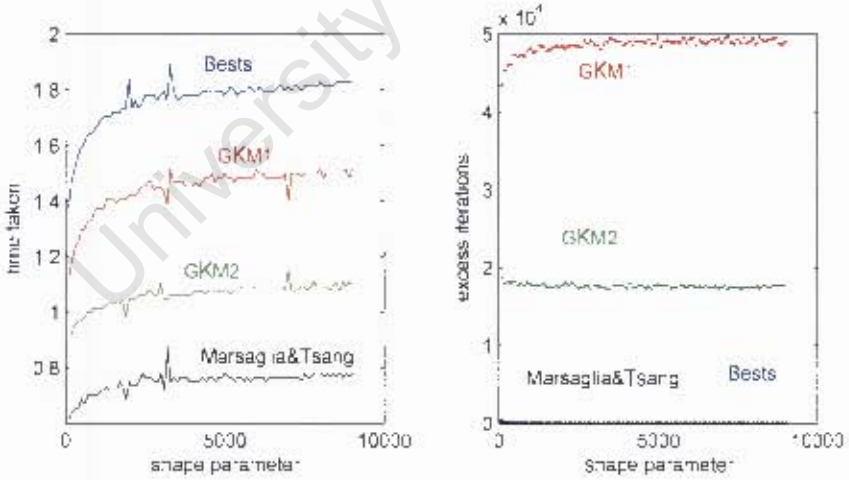

FIGURE 6.4: Figures showing the time taken (left) and number of excess iterations (right) for the various sampling techniques where the shape parameter is large ($\gg 1$)

due to *Marsaglia & Tsang* [63] seems to be the best. It offers an almost constant generation time across all shape parameters and this is less than 1 second for 50000 draws. It also offers low rejection rates. There are very small regions on the unit interval, where the other methods offer a lower rejection rate, but it should also be noted that in these regions the generators may not be all that accurate. In fact, for shape parameters less than $\alpha = 0.06$ the Marsaglia & Tsang method was found to return inaccurate values (mostly zeros), while Johnk's algorithm produced unreliable results for $\alpha < 0.07$. This posed a problem for the bridge sampling algo-

rithms, where small shape parameters were required. To address this, the method of L'Ecuyer and Simard (2006) [56] was implemented. For sampling from the gamma distribution with small shape parameters, it is recommended that an alternative approximation technique be used.

The GKM algorithms did not yield accurate results (Table 6.3). It can be seen from the shape of their graph in Figure 6.4 that they are not at all efficient for large α. The GKM1 required an additional 49247 loops just to generate 50000 random variables, corresponding to an acceptance rate of around 50%. Algorithm GKM2 fared slightly better on the acceptance rate side, but its moments were still fairly inaccurate, even after 60000 variates were generated. These two algorithms were not pursued any further. Best's algorithm has a rejection rate which exhibited

STATISTIC	**Expected**	**Ahrens-Dieter**	**Marsaglia & Tsang**	**Johnck**
mean	0.2500	0.2480	0.2518	0.2505
variance	0.1250	0.1215	0.1285	0.1263
skewness	2.8284	2.8489	2.9545	2.9400
kurtosis	12.0000	12.6907	13.7524	14.2227
time (s)	–	1.2030	0.8750	1.1100

TABLE 6.1: Table showing expected and observed statistics from simulation of 60000 gamma variates with shape parameter less than 1

an exponentially[1] decreasing relationship with an increase in shape parameter. It is better to use alternative algorithms for shape parameters in the region of 1 as the acceptance rate here is also in the region of 50%. Their use becomes more feasible as α increases.

The sample moments of generated variates were also calculated, and these were compared to empirical moments of the gamma distribution for the associated parameters. A gamma distribution has the following first four central (and normalised) moments:

$$\mu = \frac{\alpha}{\beta} \qquad \sigma^2 = \frac{\alpha}{\beta^2}$$

$$\varsigma = \frac{2}{\alpha} \qquad \kappa = \frac{6}{\alpha}$$

Moments from the simulation of 60000 gamma variates using the different methods have been

	Bests		**Marsaglia & Tsang**	
	observed	expected	observed	expected
mean	16.7009	16.6667	16.6686	16.6667
variance	65.0364	55.5556	55.3912	55.5556
skewness	1.0266	0.8944	0.8812	0.8944
kurtosis	1.8222	1.2000	1.2040	1.2000
time (s)	2.2970		0.8750	

TABLE 6.2: Table showing expected and observed statistics from simulation of 60000 gamma variates with shape parameter greater than 1

recorded in three different tables. Table 6.1 contains a summary of the statistics with $\alpha = 0.5, \beta = 2$. Table 6.2 has the results for the case where $\alpha = 5, \beta = 0.3$. Finally Table 6.3 contains a summary of the moments for the GKM algorithms, where the numbers can be seen to be inaccurate. The reason for this could not be established, even after careful inspection of the

[1] by fitting an exponential curve to the data a relationship with an $R^2 = 0.914$ was obtained

	GKM1		GKM2	
	observed	expected	observed	expected
mean	11.3620	16.6667	12.9531	16.6667
variance	40.1195	55.5556	35.5514	55.5556
skewness	0.6545	0.8944	0.5639	0.8944
kurtosis	0.6147	1.2000	1.1215	1.2000
time (s)	1.5620		1.656	

TABLE 6.3: Table showing expected and observed statistics from simulation of 60000 gamma variates for the GKM algorithms

code. The algorithm of choice was thus the method due to *Marsaglia & Tsang* [63] which, as subsequently discovered (and mentioned in Chapter 4) is the method implemented in Matlab.

6.2 Simulation of inverse Gaussian random variables

	Observed	Expected	Observed	Expected	Observed	Expected
(μ, λ)	$(0.2, 16)$		$(7, 10)$		$(0.3, 0.6)$	
mean	0.2002	0.2000	7.0221	7.0000	0.3006	0.3000
variance	0.0005	0.0005	34.4638	34.3000	0.0460	0.0450
skewness	0.3439	0.3354	2.4800	2.5100	2.1475	2.1213
kurtosis	0.1679	0.1875	9.9404	10.5000	7.5065	7.5000
time	0.4530		0.4680		0.4530	

TABLE 6.4: Summary of results for generation of 60000 IG random variables using the method of Michael, Schucany and Haas (1976) for various combinations of μ and λ

A similar process was undertaken for simulation of inverse Gaussian random variables. Only one algorithm was available which turned out to be accurate and fast. The results are summarised in Table 6.4.

6.3 Simulating beta random variables

Two methods for sampling from the beta distribution were outlined in Chapter 4. One uses a ratio of two gamma variables and thus depends on the quality of the gamma variables used. The other method is an accurate inverse transform method using series representations for the cumulative symmetric beta distribution function. Using a ratio of gammas method (Proposition 4.1) in Matlab, it took 6 seconds to generate 50000 beta (3,3) variables. Using the BETASYMINV function to generate the same number of variables took a third of the time, with similar accuracy, illustrating the power of using an efficient inverse transform method over rejection based sampling. The shortfall, however, is that this method is only useful for symmetric beta inversion. Sampling from a beta distribution where $\alpha \neq \beta$ requires the use of an alternative method (such as ratio of gammas). An alternative approach is to make use of the BETARND function in Matlab, which according to the documentation simply uses the ratio-of-gammas method.

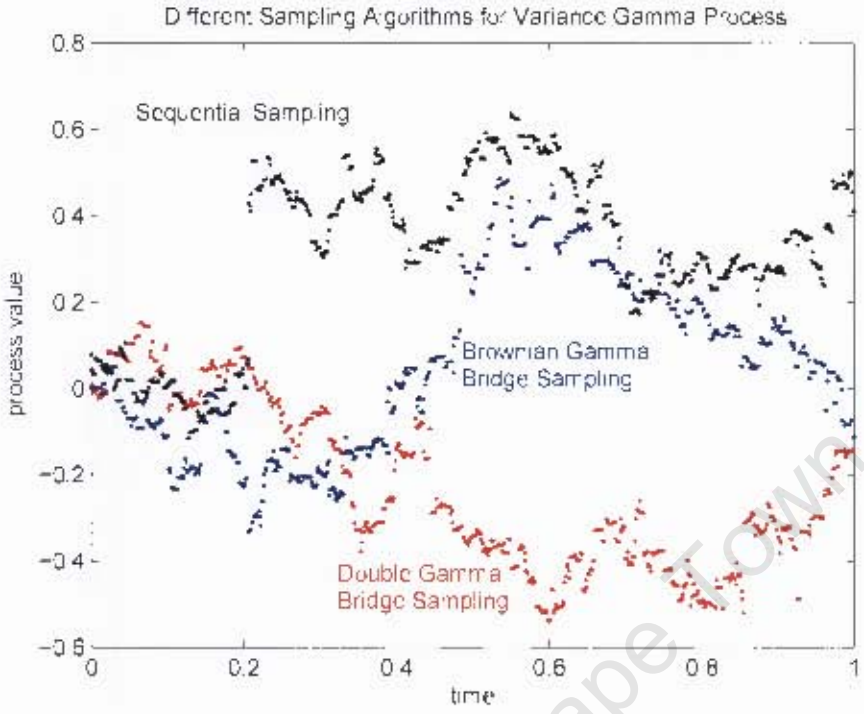

FIGURE 6.5: Graph illustrating variance-gamma processes with identical parameters, simulated using the three algorithms (Section 5.4)

6.4 Simulation of Variance-Gamma sample paths

Variance-gamma process sample paths were generated for a varying number of simulations. This number ranged from 5000 sample paths up to 100000. Each sample path contained 1025 points in it ($2^{10}+1$). Although one could speed up the moment calculation by only sampling the ter-

DGBS	Mean	Variance	Skewness	Kurtosis
5000	-0.0491	0.1257	-0.0172	3.2015
10000	-0.0445	0.1268	-0.0452	3.3243
20000	-0.0525	0.1266	-0.0343	3.2919
50000	-0.0517	0.1239	-0.0256	3.2889
100000	-0.0490	0.1256	-0.0489	3.3194
Expected	-0.0502	0.1257	-0.0423	3.3048

TABLE 6.5: Sample moments at a fixed point on variance-gamma sample paths using double gamma bridge sampling

minal value of the process (at maturity), it was decided to simulate the entire path, to assist in comparing which simulation algorithm is quicker and more efficient for generating entire sample paths. This also meant that expected moments could be calculated at arbitrary points on the sample path. Calculating values at maturity differs from the other calculations on the sample path and this would determine which algorithm was faster or slower. The parameters used in the simulation of the variance-gamma processes were $\theta = -0.1, \sigma = 0.4, \kappa = 0.2, T = 1$.

It is interesting (but most likely coincidental) that the mean and variance seemed to display exceptional convergence for simulation of 5000 sample paths. The skewness and kurtosis were

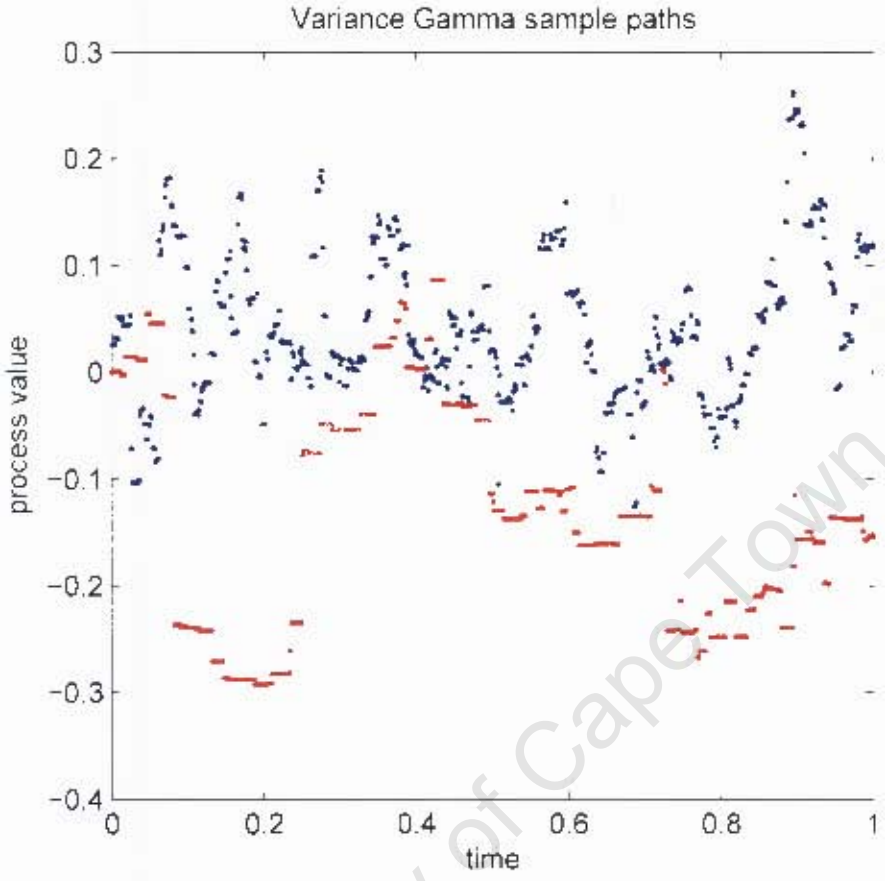

FIGURE 6.6: Plot showing how the fundamental shape of the variance-gamma process changes when κ is varied

a little bit off, but all four moments did tend to converge towards their expected values as the number of simulations was increased.

Figure 6.5 shows sample paths of the variance-gamma process using the three different sampling techniques. The simulation was carried out using the same parameters. From this the similar pathwise nature of the sample paths can be seen. On inspection of Tables 6.5 to 6.7 one can more or less gauge that convergence is of a typical Monte Carlo order, with at least 50000 simulations required before decent convergence is attained. This was true for all sampling procedures – there does not appear to be one which offers superior convergence. It may be possible to improve on this through the use of Quasi Monte Carlo techniques. Pathwise construction using the Sobol' numbers was attempted, but only for the Brownian gamma bridge and sequential sampling methods. In doing this, however, the paths no longer converged to the numbers they were supposed to. It is not immediately obvious, from the theory, how low discrepancy sequences should be applied in order to achieve this, and if it is not done correctly then attempting their use is logically equivalent to a waste of time.

From Figure 6.6 one can see the effect of decreasing the κ parameter. As mentioned in the theory of the variance-gamma process, one is able to recover the normal density in the limit $\kappa \to 0$. One can see the sample path for small κ (blue) displaying greater variation and smaller jumps more often, characteristic of Brownian motion, although this is by no means an accurate

BGBS	Mean	Variance	Skewness	Kurtosis
5000	-0.0504	0.0810	-0.1244	4.2103
10000	-0.0497	0.0838	-0.2773	4.3473
20000	-0.0474	0.0811	-0.1658	4.2740
50000	-0.0504	0.0818	-0.2282	4.1996
100000	-0.493	0.0809	-0.1857	4.1822
Expected	-0.0500	0.0810	-0.2100	4.3471

TABLE 6.6: Sample moments at a fixed point on variance-gamma sample paths using Brownian gamma bridge sampling

SEQUENTIAL	Mean	Variance	Skewness	Kurtosis
5000	0.0504	0.0794	0.2625	4.4074
10000	0.0537	0.0841	0.4017	5.1160
20000	0.0505	0.0812	0.3168	5.3728
50000	0.0498	0.0807	0.3500	4.8506
100000	0.0491	0.0814	0.3031	4.8473
Expected	0.0500	0.0815	0.3153	5.1276

TABLE 6.7: Sample moments at a fixed point on variance-gamma sample paths using sequential sampling

assessment of whether the process is a Brownian motion or not.

Table 6.8 has a list of the times taken for a total number of sample paths to be generated ac-

process	VARIANCE-GAMMA			NORMAL INVERSE GAUSSIAN
nsims	DGBS	BGBS	SEQUENTIAL	SEQUENTIAL
5000	685	634	120	125
10000	1359	1269	226	243
20000	2749	2594	608	488
50000	7125	6496	1689	1243
100000	13900	13127	3513	2573

TABLE 6.8: Times (in seconds) to simulate a given number of sample paths for specific simulation algorithms

cording to the various algorithms for both the variance-gamma and the normal inverse Gaussian processes. The sequential sampling algorithms appear to be the quickest, while for the bridge algorithms, the Brownian gamma bridge sampling algorithm is slightly faster than the double gamma bridge sampling algorithm. This may be due to the fact that the double gamma algorithm has to draw two gamma variates, which is slower than drawing 1 gamma and 1 normal variate in Matlab. Also, the reason for the difference in the expected values of the parameters in Table 6.6 is simply because the central moments were calculated at different points along the sample paths.

6.5 Simulation of NIG sample paths

For the normal inverse Gaussian process, only a sequential sampling algorithm was available. Thus, it was tested in the same way the variance-gamma process was. Sample paths are shown

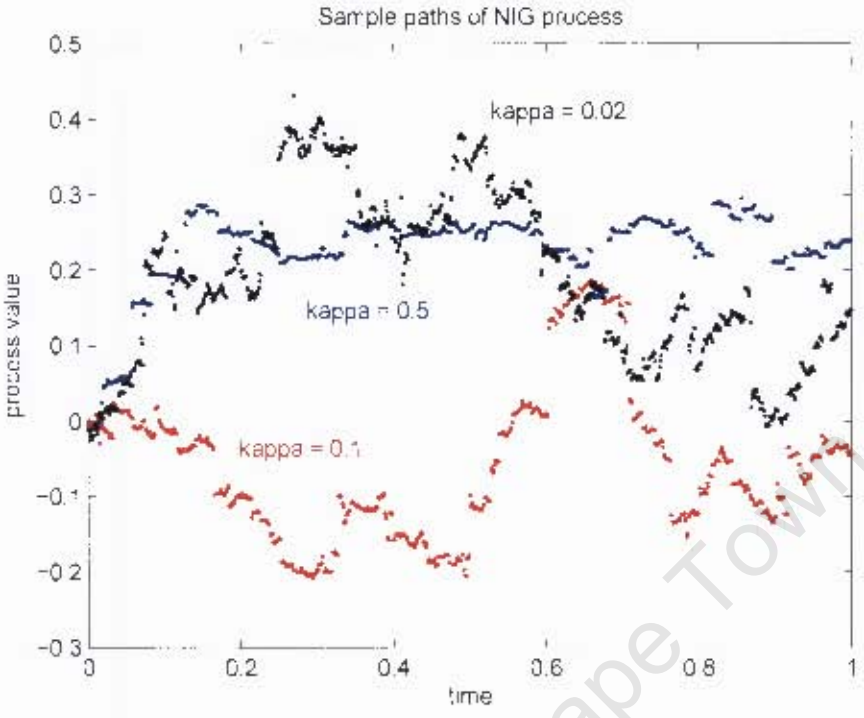

FIGURE 6.7: Typical sample paths for the normal inverse Gaussian process

in Figure 6.7. The effect of varying the κ parameter can be seen by looking at Figure 6.7. This controls the frequency of jumps present in the process. For large κ the sample paths seem to tend towards a piecewise constant function with large jumps. $\kappa \to 0$ seems to produce paths with smaller but more frequent jumps. Notice, however, that the small κ limit produces sample paths which are not entirely equivalent to the variance-gamma process. In this process, although most of the jumps are small, every now and then there are jumps which are significantly large. This process, as $\kappa \to 0$ does not seem to converge towards a Brownian motion, as can be verified theoretically.

SEQUENTIAL	Mean	Variance	Skewness	Kurtosis
5000	-0.0522	0.1263	-0.0499	3.232
10000	-0.0459	0.1251	-0.0685	3.271
20000	-0.0479	0.1258	-0.0358	3.281
50000	-0.0493	0.1256	-0.0638	3.298
100000	-0.0495	0.1255	-0.0460	3.288
Expected	-0.0500	0.1253	-0.0424	3.306

TABLE 6.9: Sample moments at a fixed point on the sample paths of NIG processes

6.6 Implementation of the CPP Lévy process approximation

The simulation of various types of processes was attempted using the compound Poisson approximation. This included the variance-gamma process and the normal inverse Gaussian process. Figure 6.8 shows generated sample paths for the normal inverse Gaussian and variance-gamma

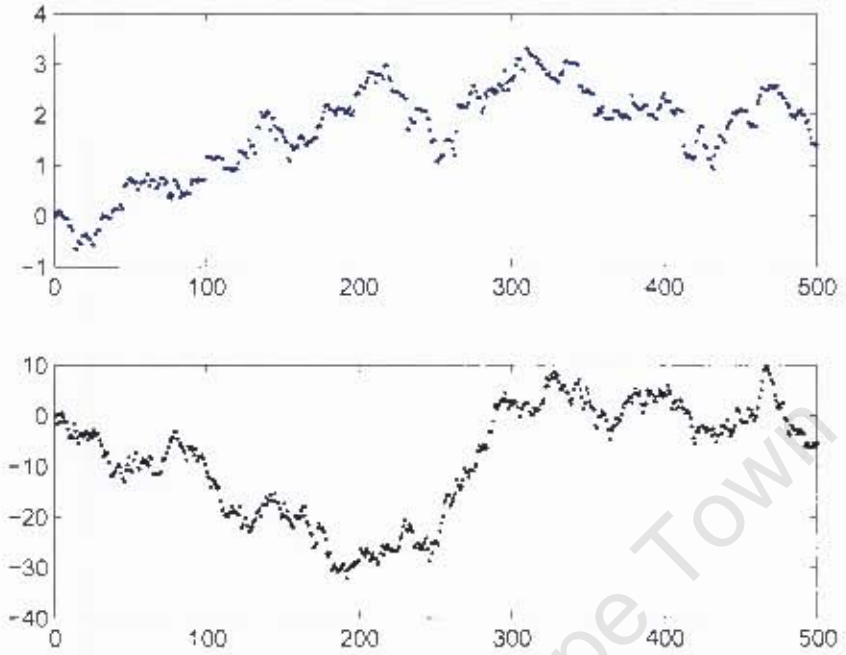

FIGURE 6.8: Sample paths of the VG (top) and NIG (bottom) processes using a Compound Poisson approximation

processes. Note that the bridge and sequential sampling methods were carried out on a much finer time grid. In modelling, since the time variable is rather arbitrary, it is a quantity which will be scaled or calibrated according to the problem at hand. The model parameters are determined by the data and the frequency at which it has been sampled.

With implementing these approximations, the first task involved calculating graphs of the Lévy measure. These were then discretised according to the number of compound Poisson processes that were specified along with the starting points (or boundary values) of the bins. Romberg integration was then used to determine the parameters for each bin so that jump times could be simulated for each Poisson process. Once the jump times on a time interval $[0, T]$ were obtained, the value of the process could be evaluated for arbitrary times between 0 and T using the jump size parameters (c_i). The resulting sample path could then be plotted. All that was required was the expression for the Lévy measure, values for the associated parameters and some cutoff time, T.

Although the simulation process involved is quite generic, it is not without its issues. Firstly, there is plenty of room for error if the code is not thoroughly checked and implemented with caution. This is because at any given point in the execution of the algorithm, there are quite a lot of things to keep track of: simulation of all the jump times for the d Poisson processes, checking how many times each process had jumped at each particular time point and calculation of the λ_i and c_i parameters for each Poisson process amongst others. Also, in the event of the use of an approximating Brownian motion for the small jumps, one has to ensure that the variance used for this process has been calculated correctly. The intricacy and sheer volume of calculations which must be carried out for this approximation algorithm slow the algorithms down substantially, which can also be a problem, particularly if their accuracy is under question.

Another issue with the approximation approach, is that it is more difficult to verify whether the generated paths are actually accurate. For instance, by looking at Figure 6.8 it is not clear whether the process is simply missing the large jumps or if they have not been captured by the use of multiple Poisson processes. It is not at all scientific to claim that two processes are similar by looking at their graphs, but there does not appear to be much literature available on determining the accuracy of a generated process. Due to the buggy nature of this process, it is recommended that it be used with caution as mistakes could creep into the model very easily, rendering the results inaccurate.

The author did not have much success with the implementation of the low-discrepancy sequences or Quasi-Monte Carlo techniques, as was alluded to earlier. They were applied to the bridge sampling algorithms by using 1 Sobol' number per sample path to obtain the terminal value of the process, with the rest of the sample path calculated using psuedo-random numbers. As mentioned, this exacerbated the convergence of the process, which is far from the objective of quasi-random numbers. Other variance reduction techniques mentioned in the literature may also be considered. These include stratified sampling, latin hypercube sampling, control variates and antithetic variates. Control variates can be used if one knows the payoff structure of an option being priced and are thus not useful to reduce variance in simple sample paths. Antithetic variates make use of the fact that if U_1 is uniformly distributed on [0,1], then so is $1 - U_1$. Again, these are most easily applied to the subordinated Brownian motion examples and could possibly be incorporated into the less straightforward simulation algorithms but this has not been considered in this dissertation.

6.7 Concluding remarks and further research recommendations

The simulation of Lévy processes is quite an involved and tricky science, requiring knowledge of a diverse range of simulation techniques and distributional properties. This dissertation has looked in detail at the simulation algorithms of the more popular Lévy processes: the variance-gamma process, the normal inverse Gaussian process and jump-diffusion processes. It has also attempted to provide a methodology for simulating more general Lévy processes although the accuracy of this algorithm has not been tested. Since price processes appear to move by jumps, it seems natural to select, as a model, something which incorporates jumps as well. This dissertation has aimed to introduce Lévy processes at a broader level, to educate and assist in the decision making process of which Lévy process could be used in a particular scenario.

Research of the ideas undertaken here could be further investigated in a multivariate (or multi-dimensional) setting. This would be useful when pricing baskets of options or more exotic derivative structures, where interdependence between the price processes would play an important role. For this the use of Lévy copulas has been advocated, albeit at a very shallow level. This choice offers flexibility with regard to modelling dependent Lévy processes. Also, this dissertation has laid a groundwork for the simulation of Lévy processes. These methods could be applied to a calibration and pricing framework, with the hope that the algorithms proposed here can be used or discarded with confidence in calibrating and pricing problems. These are issues that have not been dealt with at all, but are obviously key in the financial industry. Adopting a specific calibration method to obtain a risk-neutral Lévy process would lead to prices for options which are of practical use to the finance industry. In addition to these, the notion of using Quasi-Monte Carlo methods or variance reduction in obtaining sample paths could be explored as this section was not given much attention.

There is certain merit to be gained from the use of Lévy processes, and it seems naïve to write

them off simply because they are too complex. Although they seem to live in the shadow of the Brownian motion models, they are powerful and have the ability to capture more of the empirical properties present in asset returns. One would hope that this would translate into more accurate prices, but this is difficult to verify. There does not appear to be much consensus on jump models in international markets, but that does not mean that suitable and valuable frameworks for the evaluation of option structures cannot be built by using Lévy processes which look set to still have an important role to play into the future.

Bibliography

[1] *Numerical Recipes: The Art of Scientific Computing*, Third ed. Cambridge University Press, 2007.

[2] ABRAMOWITZ, M., AND STEGUN, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth Dover printing, tenth GPO printing ed. Dover, New York, 1964.

[3] ACKLAM, P. http://home.online.no/~pjacklam/notes/invnorm/ (Accessed March 2008).

[4] AHRENS, J., AND DIETER, U. Computer methods for sampling from gamma, beta, Poisson and binomial distributions. *Computing 25* (1974), 193–208.

[5] AKGIRAY, V., AND BOOTH, G. The stable law model of stock returns. *Journal of Business Economic Statistics 6* (1988), 51–57.

[6] ANTONOV, I., AND SALEEV, V. An economic method of computing LP_τ sequences. *U.S.S.R Computing Maths. Math. Phys. 19* (1979), 252–256.

[7] ASMUSSEN, S., AND GLYNN, P. W. *Stochastic Simulation*. Springer Science and Business Media, LLC, 2007.

[8] BARNDORFF-NIELSEN, O. E. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics 5* (1978), 151–157.

[9] BARNDORFF-NIELSEN, O. E. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics / Blackwell Publishers 24* (1997), 1–13.

[10] BARNDORFF-NIELSEN, O. E. Normal inverse Gaussian distributions and the modelling of stock returns. Department of Theoretical Statistics, Aarhaus University, 1997. Research Report No. 300.

[11] BARNDORFF-NIELSEN, O. E., AND HALGREEN, C. Infinite divisibility of the hyperbolic and generalised inverse Gaussian distributions. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebeite 38* (1977), 309–311.

[12] BECKER, R. I. *Financial Mathematics in Excel (course notes)*. 2006.

[13] BEST, D. A note on gamma variate generators with shape parameter less than unity. *Computing 30* (1983), 185–188.

[14] BIANE, P., PITMAN, J., AND YOR, M. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bulletin of the American Mathematical Society 38* (2001), 435–465.

[15] BLACK, F., AND SCHOLES, M. The pricing of options and corporate liabilities. *The Journal of Political Economy 81* (1973), 637–654.

[16] BLATTBERG, R., AND GONEDES, N. A comparison of the stable Paretian and Student t distributions as statistical models for prices. *Journal of Business 47* (1974), 244–280.

[17] BOX, G. E. P., AND MULLER, M. E. A note on the generation of random normal deviates. *Annals of Mathematical Statistics 29* (1958), 610–611.

[18] BREEDEN, D. T., AND LITZENBERGER, R. H. Prices of state-contingent claims implicit in option prices. *Journal of Business 51*, 4 (1976), 621–51.

[19] CARR, P., CHANG, E., AND MADAN, D. The variance gamma process and option pricing. *European Finance Review 2* (1998), 79–105.

[20] CARR, P., GEMAN, H., MADAN, D., AND YOR, M. The fine structure of asset returns: An empirical investigation. *Journal of Business 75* (2002), 305–332.

[21] CARR, P., AND MADAN, D. B. Option valuation using the fast fourier transform. *Journal of Computational Finance 2* (1999), 61–73.

[22] CASTAGNA, A., AND MERCURIO, F. The Vanna-Volga Method for Implied Volatilities. *Risk* (2007), 106–111.

[23] CHAMBERS, J., MALLOWS, C., AND STUCK, B. A method for simulating stable random variables. *Journal of American Statistical Association 71*, 340–344.

[24] CHENG, R., AND FEAST, G. M. Some simple gamma variate generators. *Applied Statistics 28* (1979), 290–295.

[25] CONT, R., AND TANKOV, P. *Financial Modelling with Jump Processes.* Chapman & Hall CRC Financial Mathematics Series, 2004.

[26] DACOROGNA, M., MÜLLER, U., AND PICTET, O. Heavy Tails in High-Frequency Financial Data. *A Practical Guide to Heavy Tails* (1998), 55–79.

[27] DEVROYE, L. *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986.

[28] DUPIRE, B. Pricing with a smile. *Risk 7* (1994), 18–20.

[29] EBERLEIN, E., AND KELLER, U. Hyperbolic distributions in finance. *Bernoulli 1* (1995), 281–299.

[30] EBERLEIN, E., KELLER, U., AND PRAUSE, K. New insights into smile, mispricing and value at risk: The hyperbolic model. *Journal of Business 71* (1998), 371–406.

[31] EBERLEIN, E., AND PRAUSE, K. The generalised hyperbolic model: Financial derivatives and risk measures. FDM Preprint 56, University of Freiburg, 1998.

[32] EBERLEIN, E., AND RAIBLE, S. Term structure models driven by general Lévy processes. *Mathematical Finance 9* (1999), 31–53.

[33] FAMA, E. The behaviour of stock market prices. *Journal of Business 38* (1965), 34–105.

[34] FAURE, H. Discrépence de suites associées à un système de numération (en dimension s). *Acta Arithmetica 41* (1982), 337–351.

[35] FELLER, W. *An Introduction to Probability Theory and its Applications, Volume I*, third ed. Wiley, 1968.

[36] FISHMAN, G. *Monte Carlo: Concepts, algorithms and applications.* Springer-Verlag, New York, 1996.

[37] GEMANN, H. Pure jump Lévy processes for asset price modelling. *Journal of Banking & Finance 26* (2002), 1297–1316.

[38] GEMANN, H., MADAN, D., AND YOR, M. Asset Prices are Brownian Motion: Only in Business Time. 1998.

[39] GENTLE, J. E. *Random Number Generation and Monte Carlo Methods.* Springer-Verlag, New York, 1998.

[40] GLASSERMAN, P. *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, New York, 2004.

[41] GOOD, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika 40* (1953), 237–260.

[42] GRIGELIONIS, B. Processes of Meixner type. *Lithuanian Mathematics Journal 39*, 1 (2000), 33–41.

[43] HEATH, D., JARROW, R., AND MORTON, A. Bond Pricing and the Term Structure of Interest Rates: A New Methodology. *Econometrica 60* (1992), 77–105.

[44] HÖRMANN, W., AND DERFLINGER, G. The transformed rejection method for generating random variables, an alternative to the ratio of uniforms method. *Communications in Statistics - Simulation and Computation 23*, 847–860.

[45] HOUGAARD, P. Survival models for heterogeneous populations derived from stable distributions. *Biometrika 73* (1986), 387–396.

[46] HULL, J. *Options, Futures and other Derivatives*, 6th ed. Prentice Hall, New Jersey, 2006.

[47] HULL, J., AND WHITE, A. The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance 42* (1987), 281–300.

[48] HUNT, P., AND KENNEDY, J. *Financial Derivatives in Theory and Practise.* John Wiley & Sons, 2004.

[49] JÄCKEL, P. *Monte Carlo Methods in Finance.* John Wiley & Sons, Ltd, 2002.

[50] JANG, J., AND GALANTI, B. Low discrepancy sequences: Monte Carlo simulation of option prices. *Journal of Derivatives* (1997), 63–83.

[51] JANSEN, D., AND DE VRIES, C. On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *Review of Economics and Statistics 73* (1991), 18–24.

[52] JÖHNCK, M. Erzeugung von Betaverteilter und Gammaverteilter Zufallszahlen. *Metrika 8* (1964), 5–15.

[53] JØRGENSEN, B. *Statistical Properties of the Generalised Inverse Gaussian Distribution. Lecture Notes in Statistics.* Springer, 1982.

[54] KACENAK, M. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=34, & objectType=file (Accessed January 2008).

[55] KINDERMANN, A., AND MONAHAN, J. Computer generation of random variables using the ratio of uniforms deviates. *ACM Transaction on Mathematical Software 3* (1977), 257–260.

[56] L'ECUYER, P., AND SIMARD, R. Inverting the symmetrical beta distribution. *ACM Trans. Math. Softw. 32*, 4 (2006), 509–520.

[57] LONGIN, F. The Asymptotic Distribution of Extreme Stock Market Returns. *Journal of Business 69* (1996), 383–408.

[58] MADAN, D. Purely discontinuous asset prices (working paper). 1999.

[59] MADAN, D., AND MILNE, F. Option pricing with v.g. martingale components. *Journal of Mathematical Finance 1* (1991), 39–55.

[60] MANDELBROT, B. The variation of certain speculative prices. *Journal of Business 36* (1963), 394–419.

[61] MARSAGLIA, G. The exact-approximation method for generating random variables in a computer. *Journal of the American Statistical Association 79* (1977), 218–221.

[62] MARSAGLIA, G. The squeeze method for generating gamma variates. *Computers and Mathematics with Applications 3* (1977), 321–325.

[63] MARSAGLIA, G., AND TSANG, W. A simple method for generating gamma variates. *ACM Transactions on Mathematical Software 26*, 3 (2000).

[64] MARSAGLIA, G., AND TSANG, W. The ziggurat method for generating random variables. *Journal of Statistical Software* (2000).

[65] MATSUMOTO, M., AND NISHIMURA, T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulations: Special Issue on Uniform Random Number Generation* (1998).

[66] MCLEISH, D. L. *Monte Carlo Simulation and Finance.* John Wiley & Sons, 2005.

[67] MERTON, R. Option pricing when underlying stock returns are discontinuous. *J, Financial Economics 3* (1976), 125–144.

[68] MICHAEL, J., SCHUCANY, W., AND HAAS, R. Generating random variates using transforms with multiple roots. *American Statistician 30* (1976), 88–90.

[69] MUSIELA, M., AND RUTKOWSKI, M. Continuous-time term structure models: Forward measure approach. *Finance and Stochastics 1* (1997), 261–291.

[70] NIEDERREITER, H. Random Number Generation and Quasi-Monte Carlo Methods. *Society for Industrial and Applied Mathematics* (1992).

[71] NOLAN, J. Stable Distributions: Models for Heavy Tailed Data. http://academic2.american.edu/~jpnolan/stable/chap1.pdf, Mathematics/Statistics Department, American University (Accessed February 2008), 2005.

[72] RYDBERG, T. H. The normal inverse gaussian process: Simulation and approximation. *Comm. Stat.: Stoch Models 13* (1997), 887–910.

[73] SAMUELSON, P. A rational theory of warrant pricing. *Industrial Management Review 6* (1965), 13–31.

[74] SATO, K. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge, 1999.

[75] SCHOUTENS, W. The Meixner Process: Theory and Applications in Finance. *Mini Proceedings: 2nd MaPhySto Conference on Lévy Processes - Theory and Applications* (2002).

[76] SCHOUTENS, W. *Lévy processes in Finance.* John Wiley & Sons, Ltd, 2003.

[77] SCHOUTENS, W., AND TEUGELS, J. L. Lévy processes, polynomials and martingales. *Communications in Statistics: Stochastic Models 14* (1998), 335–349.

[78] SOBOL', I. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Journal of Computational Mathematics and Mathematical Physics 7* (1967), 86–112.

[79] TANKOV, P. *Levy Processes in Finance: Inverse Problems and Dependence Modelling.* PhD thesis, 2004.

[80] TWEEDIE, M. An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference* (1984), J. Ghosh and J. Roy, Eds.

[81] VASICEK, O. An Equilibrium characterization of the Term Structure. *Journal of Financial Economics 5* (1977), 177–188.

[82] WALLACE, C. Transformed rejection generators for gamma and normal psuedo-random variables. *Australian Computer Journal 8* (1976), 103–105.

[83] WEST, G. Calibration of the SABR model in Illiquid Markets. *Applied Mathematical Finance 12* (2005), 371–385.